

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26 2005, 13:45:55 ; Search time 44 Seconds
(without alignments)
900.878 Million cell updates/sec

Title: US-10-721-553-2
Perfect score: 2764
Sequence: 1 MAPIQTQQRDEGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/protdata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/protdata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/protdata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/protdata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/protdata/1/iaa/PCITUS-COMB.pep.*
6: /cgn2_6/protdata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764	100.0	531	2	US-08-933-750C-9
2	2764	100.0	531	3	US-09-234-613-9
3	2764	100.0	531	4	US-09-647-143-2
4	595	21.5	115	4	US-09-513-999C-7407
5	266	9.6	481	4	US-09-248-796A-18683
6	234.5	8.5	1742	4	US-09-386-962C-4
7	234.5	8.5	1742	4	US-09-386-959-4
8	233.5	8.4	930	4	US-09-200-650B-3
9	226	8.2	918	4	US-09-200-650B-1
10	225.5	8.2	933	3	US-08-293-728-2
11	225.5	8.2	936	3	US-09-421-868-2
12	225.5	8.2	936	4	US-08-956-171B-5249
13	225.5	8.2	936	4	US-08-781-986A-5249
14	223.5	8.1	1315	4	US-09-200-650B-5
15	220.5	8.0	1259	4	US-09-949-016-10366
16	218	7.9	1166	4	US-09-200-650B-7
17	212.5	7.7	287	4	US-09-710-279-468
18	212.5	7.7	1092	4	US-09-147-405B-15
19	203.5	7.4	414	4	US-09-248-796A-19046
20	203	7.3	257	3	US-09-461-697-188
21	203	7.3	272	3	US-09-461-697-186
22	199.5	7.2	599	4	US-09-538-092-864
23	198.5	7.2	238	3	US-09-461-697-190
24	196	7.1	781	4	US-09-949-016-9773
25	195	7.1	231	3	US-09-461-697-194
26	195	7.1	232	3	US-09-461-697-192
27	195	7.1	764	4	US-09-538-092-944

28 191 6.9 598 4 US-09-538-092-1083
29 189 6.8 764 4 US-09-370-838-67
30 189 6.8 764 4 US-09-854-133-67
31 186 6.7 3135 1 US-08-323-170B-2
32 186 6.7 3135 3 US-08-954-441-2
33 185.5 6.7 1162 2 US-08-728-323A-2
34 185.5 6.7 1162 3 US-09-298-568-2
35 185.5 6.7 1162 4 US-09-410-393-2
36 185.5 6.7 1162 4 US-09-894-273-2
37 185 6.7 402 4 US-09-248-796A-18910
38 184 6.7 674 3 US-08-893-852A-1
39 183 6.6 486 4 US-09-710-279-788
40 180 6.5 487 4 US-09-386-962C-14
41 180 6.5 487 4 US-09-386-953-65
42 178.5 6.5 40 4 US-09-647-143-16
43 178.5 6.5 1269 4 US-09-949-016-7349
44 178.5 6.5 1269 4 US-09-949-016-7350
45 178 6.4 1444 4 US-09-949-016-9652

ALIGNMENTS

RESULT 1
US-08-933-750C-9
; Sequence 9, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PITUNOR01

102c

```
; CLONE: 98974
US-08-933-750C-9

Query Match      100.0%; Score 2764; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYFPDQNRVQ 60
DB 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYFPDQNRVQ 60
QY 61 YKATSLKQKHGDLLETPDLGVTTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHGDLLETPDLGVTTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
QY 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQFTBEEIYKDRDSQITAE 180
DB 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQFTBEEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDKMWINPCAQVIFDSDPAPKOTSGAAALEM 240
DB 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDKMWINPCAQVIFDSDPAPKOTSGAAALEM 240
QY 241 MSQAMIRGMDREGNOFVAYFLPVEETLKRRKQDEEEMDYAPDDVDYKIAREYNWVK 300
DB 241 MSQAMIRGMDREGNOFVAYFLPVEETLKRRKQDEEEMDYAPDDVDYKIAREYNWVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKSGSSEKSGSEDEHSGSESER 420
DB 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKSGSSEKSGSEDEHSGSESER 420
QY 421 EGGDRDEASDKSGGDESEDEARAAKDEEIFGSDADSDDDADSDDEDRGQAQGGSDN 480
DB 421 EGGDRDEASDKSGGDESEDEARAAKDEEIFGSDADSDDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGNGGGQGRSHRSASPPFSGSEHSAQEDGSEAAASDSSDSDSD 531
DB 481 DSDSGNGGGQGRSHRSASPPFSGSEHSAQEDGSEAAASDSSDSDSD 531

RESULT 3
US-09-647-143-2
; Sequence 2, Application US/09647143
; Patent No. 6680196
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; APPLICANT: University of Nebraska Board of Regents

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PITUNOR01
; CLONE: 98974
US-09-234-613-9

Query Match      100.0%; Score 2764; DB 3; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYFPDQNRVQ 60
DB 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYFPDQNRVQ 60
QY 61 YKATSLKQKHGDLLETPDLGVTTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHGDLLETPDLGVTTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
QY 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQFTBEEIYKDRDSQITAE 180
DB 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQFTBEEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDKMWINPCAQVIFDSDPAPKOTSGAAALEM 240
DB 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDKMWINPCAQVIFDSDPAPKOTSGAAALEM 240
QY 241 MSQAMIRGMDREGNOFVAYFLPVEETLKRRKQDEEEMDYAPDDVDYKIAREYNWVK 300
DB 241 MSQAMIRGMDREGNOFVAYFLPVEETLKRRKQDEEEMDYAPDDVDYKIAREYNWVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKSGSSEKSGSEDEHSGSESER 420
DB 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKSGSSEKSGSEDEHSGSESER 420
QY 421 EGGDRDEASDKSGGDESEDEARAAKDEEIFGSDADSDDDADSDDEDRGQAQGGSDN 480
DB 421 EGGDRDEASDKSGGDESEDEARAAKDEEIFGSDADSDDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGNGGGQGRSHRSASPPFSGSEHSAQEDGSEAAASDSSDSDSD 531
DB 481 DSDSGNGGGQGRSHRSASPPFSGSEHSAQEDGSEAAASDSSDSDSD 531

RESULT 3
US-09-647-143-2
; Sequence 2, Application US/09647143
; Patent No. 6680196
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; APPLICANT: University of Nebraska Board of Regents
```

```
; TITLE OF INVENTION: No. 6830196el Gene That is Amplified and
; FILE REFERENCE: UNMC63121
; CURRENT APPLICATION NUMBER: US/09/647,143
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US99/06633
; PRIOR FILING DATE: 1999-08-26
; PRIOR APPLICATION NUMBER: 50/079,649
; PRIOR FILING DATE: 1998-09-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-647-143-2

Query Match          100.0%; Score 2764; DB 4; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227; Indels 0; Gaps 0;
Matches 531; Conservative 0; Mismatches 0;

Qy 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFFDQNRVQ 60
Db 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFFDQNRVQ 60
Qy 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQQHAVVPMWRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTFEEIYKDRDSQITAIE 180
Db 121 SQQHAVVPMWRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTFEEIYKDRDSQITAIE 180
Qy 181 KTFEDAQKSIQHSYKPRVTPVENVPPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQHSYKPRVTPVENVPPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMBDEGQFVAYFLPVEETLKKGRDQEEEMDYAPDDVYDKIAREYNNVVK 300
Db 241 MSQAMIRGMBDEGQFVAYFLPVEETLKKGRDQEEEMDYAPDDVYDKIAREYNNVVK 300
Qy 301 NKASGYEENYFFIFREGDGVVYNELETRVRLSKERAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASGYEENYFFIFREGDGVVYNELETRVRLSKERAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQEARKAQLENHEPEEEEEEMETEKEAGGSDEEKEGSSSEKESSEHSGSSESR 420
Db 361 LEAQEARKAQLENHEPEEEEEEMETEKEAGGSDEEKEGSSSEKESSEHSGSSESR 420
Qy 421 EGGDRDEASDKGSGSEDESSDEARAARDKBEIFGSDADSDDEDDADSDDEDRGQAGGSDN 480
Db 421 EGGDRDEASDKGSGSEDESSDEARAARDKBEIFGSDADSDDEDDADSDDEDRGQAGGSDN 480
Qy 481 DSDSGNGGGORSRSHRSASFPFSGSHSAQEDGSEAAASDSEADSDSD 531
Db 481 DSDSGNGGGORSRSHRSASFPFSGSHSAQEDGSEAAASDSEADSDSD 531

RESULT 4
US-09-513-999C-7407
; Sequence 7407, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; PATENT NO. 6783961
; FILE REFERENCE: 59.US2.REC
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
```

```
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7407
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 25
; OTHER INFORMATION: Xaa=Glu or Gly
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 26
; OTHER INFORMATION: Xaa=Arg or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 110
; OTHER INFORMATION: Xaa=Glu or Gly
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 114
; OTHER INFORMATION: Xaa=Ala or Gly
US-09-513-999C-7407

Query Match          21.5%; Score 595; DB 4; Length 115;
Best Local Similarity 96.5%; Pred. No. 3.2e-43;
Matches 111; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFFDQNRVQ 60
Db 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFFDQNRVQ 60
Qy 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 115
Db 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 115

RESULT 5
US-09-248-796A-18683
; Sequence 18683, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18683
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18683

Query Match          9.6%; Score 266; DB 4; Length 481;
Best Local Similarity 22.0%; Pred. No. 2.8e-14;
Matches 110; Conservative 91; Mismatches 156; Indels 142; Gaps 20;

Qy 18 NSHRTL-PERSGVVCRVKYCNLSLPDIPDPKFIYFFDQNRVQYKATSL-BKQHK 71
Db 16 SSKSLKPIQDYIAKYRYTNLPPLPKFIEYNTTDPISQOGEYLISSLFKXENF 75
Qy 72 HDLLTPD--LGVTIDLINPDTY-----RIDN---VLLDPADEKLEEE 110
Db 76 QNLMERIDDLGLDLNLRNGLFSLSDKNNESVGKLYNQLHPNDRALLRDAIGIKLKN 135
Qy 111 EIQAPTSKRSQOHAKVVPWRKTEYISTEFNRYGISNEKPEVKIG-----VSVKQ 161
Db 111 EIQAPTSKRSQOHAKVVPWRKTEYISTEFNRYGISNEKPEVKIG-----VSVKQ 161
```


Db 1004 GITTTGIINGADNTIIDSFGYKPKYSGVDYVWEDTNKDGIOODNEKSGISGVKVTLKDEK 1063
QY 308 -----EENYFPIFREGD-GVYINLETRVRLSKERAKAGVQSGTNALLVVKHRDMN 357
Db 1064 GNIISTTTTIDNGKQYFNLDSGNVI IHFEKPEGWQTQTANS-----NDD 1109
QY 358 EKELEAQARKA-----QLENHEPEEEEMETEKEAGGSDEREQEKSGSSEKGSSE 410
Db 1110 EKADAGEDVRVTIHDHDFSIDNGYFDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDA 1169
QY 411 DEHSGSEERE-EGDRDEAPK-SGSGDESEDEAARAKERIFGSDADSDSDSDSDSDSDSD 468
Db 1170 DS 1229
QY 469 EDRGQAQCGSDNDSDSGNGGQGRSRSHRSRSPFSGSEHSAQEDSEAAASDS-SEAD 527
Db 1230 DSD 1286
QY 528 SDSD 531
Db 1287 SDSD 1290

RESULT 8

US-09-200-650E-3

; Sequence 3, Application US/09200650E

; Patent No. 6680195

; GENERAL INFORMATION:

; APPLICANT: Patti, Joseph M.

; APPLICANT: Foster, Timothy J.

; APPLICANT: Hook, Magnus A.O.

; APPLICANT: Eidhinn, Deirdre Ni

; APPLICANT: Perkins, Samuel L.

; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus

; FILE REFERENCE: P06283US2/BAS

; CURRENT APPLICATION NUMBER: US/09/200,650E

; CURRENT FILING DATE: 1998-11-25

; PRIOR APPLICATION NUMBER: 60/066,815

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: 60/098,427

; PRIOR FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 3

; LENGTH: 930

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-200-650E-3

Query Match 8.4%; Score 233.5; DB 4; Length 930;
Best Local Similarity 20.5%; Pred. No. 4.2e-11;
Matches 130; Conservative 77; Mismatches 211; Indels 215; Gaps 25;
QY 53 PDQNRFOYKATSELEKQKHLLTPDLGVTIDLINPDTYRIDPNVLLDPADEKLLSEEI 112
Db 298 FEQVAFARKNATIDK-----TAYKMEVT---LGNDTY-----SEEI 331
QY 113 QAPTSSKESQQAQVWPMWRKTEYISTE----- 140
Db 332 IDVYGNKKAQ-----PLISSTNYINNEDLSRMTAYVQPKNTYTKQTFVNTLTGYKFN 385
QY 141 -----FNRYGISNEK-----PEYKIGSVKQOFTBEIYKDRDSQITAEKTFDAQK 188
Db 386 PNAKNFKIYETDQNFVDSFTPDTSKLDVTDQF---DVIYSNDNKTATVD--LMKGQT 440
QY 189 SISQHSKPRVTPVEWVPFDFKMWNPQAVIFDSDPAPKDTSGAAALEMMSQAMIRG 248
Db 441 SSNKQYIIQQV-----AYPDNSSTONGKIDYILDTDKTKYSWSN-----SYSNVG 486
QY 249 MWDEEGNOFAYEL---PVEETLKKRKEDQEE-----EMD----- 280
Db 487 SSTANGDO-KKYNIGDYVWEDTNKDGQDANEKIGKGVYVILKDSNGKELDRITTTDENGK 545

QY 281 -----YAP-----DDYVD-----YKI 291
Db 546 YQFTGLSNGTYSVEFSTPAGYTPTTANVTGDDAVDSDGLTTTGVIKDADNMTLDSGYKT 605
QY 292 AR-----EYNWNVKNK-----ASKGY-----EENYFFIFREGDGV 321
Db 606 PKYSLGDYVWYDNGKDKRSTKGIKGVKVTQNEKGEVIGTTTETDENGKTRFDNLDG 665
QY 322 YNNELETRVRLSKRRRAKAGV-QSGTNALLVVKHRDMNEKELEAQARKAOLEN-HEPEEE 379
Db 666 KY-----KVIFEK---PAGLTQGTGTTTDDKADGGEVDVTTIHDHDFLTGNGYEEET 717
QY 380 EREMETEKEAGGDEEQEKSGSSEKSGSEDEHSGSEEREEREDRDEADKSGSGDEDS 439
Db 718 SD 777
QY 440 SEDEAARAKERIFGSDADSD 499
Db 778 SD 836
QY 500 ASPFPFSGSEHSAQEDGSEAAASDS-SEADSDSD 531
Db 837 DSD 869

RESULT 9

US-09-200-650E-1

; Sequence 1, Application US/09200650E

; Patent No. 6680195

; GENERAL INFORMATION:

; APPLICANT: Patti, Joseph M.

; APPLICANT: Foster, Timothy J.

; APPLICANT: Hook, Magnus A.O.

; APPLICANT: Eidhinn, Deirdre Ni

; APPLICANT: Perkins, Samuel L.

; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus

; FILE REFERENCE: P06283US2/BAS

; CURRENT APPLICATION NUMBER: US/09/200,650E

; CURRENT FILING DATE: 1998-11-25

; PRIOR APPLICATION NUMBER: 60/066,815

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: 60/098,427

; PRIOR FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 1

; LENGTH: 918

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-200-650E-1

Query Match 8.2%; Score 226; DB 4; Length 918;
Best Local Similarity 22.6%; Pred. No. 1.8e-10;
Matches 125; Conservative 75; Mismatches 208; Indels 146; Gaps 23;
QY 33 VKYCNSLPDIPP-DPKFITVFPQNRFPVQKATSELEKQKHLLTPDLGVTIDLIN--- 88
Db 276 VDYSNSNNTMPIADIK-----STNGDVAKAT-----YDILTKTFTVFTDYVNNKE 322
QY 89 -----POTYRIDPNVLLDPADEKLEEE-----EIQAPTSSKRS 121
Db 323 NINGQFSLPFTDRAPKPSGTYDANINI---ADEMFNFKITYNYVSSPIAGIDKPNGANIS 380
QY 122 QCHAKVPMWRKTEYISTEFNRYGIGSNEKEPVKIGSVKQOFTBEIYKDRDSQITAE- 180
Db 381 SOLIGVDATSGQNTYKQTVF-----VNPQRVLGNTWVYIKGYQDKI-ESSSGKVSATDT 434
QY 181 --KTPE--DAQKISQHSKPRVTPV-EVMPVPDFKMWNPQAVIFDSDPAPKDTSGA 235
Db 435 KLRIFEVNDTSKLSYADPNDSNLKEVTDQPKRIYYEHPNVASIKFGD----- 485
QY 236 AALEMMSQAMIRGMEDEGNOFVAYFLPVEETLKKRKEDQEEEMDYPDDVTDYKIAREY 295

US-08-956-171E-5249
; Sequence 5249, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 5249:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5249:
; US-08-956-171E-5249
; Query Match 8.2%; Score 225.5; DB 4; Length 936;
; Best Local Similarity 21.4%; Pred. No. 2e-10;
; Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;
; QY 83 TIDLINP--DTPRIDPNVLLDPADEKLEEEIOAPTSSKRSQQAQVVPWVRKTEYISTE 140
; DB 392 TIDQIDKNTNVR--QTIYVNPSGDNVI-----APVLT----- 422
; QY 141 FNRIGISNEKPEVKIGVSVKQQTTEEIYK-----DRDSQITAIEKTFEDAQKSIQHS 195
; DB 423 -----GNLKPTDSNALIDQQTNSIKYKVDNAADLSSEYFVNPENFEDVTNSVNTFP 476
; QY 196 KPRVTPVEVMPVFEDEKWINPACQVIFDSPPAKDTSGAAALEMM-----SQAMIRGM- 249
; DB 477 NPQYKKEFNT--EDDQITTYIVVNGHIDP-----NSKGDIALRSTLYGYNLSIWRSMS 531
; QY 250 MDEEGNQVAYFL-----PVEETLKKRKRDOEEMOYAPDDVYDYKIAREYNNVKNKA 303
; DB 532 WDNE----VAFNNGSGSDGIDKPVPPEQDPGEIEPIED-----SDS 572
; QY 304 SKGVEENTFFIFREGDGYVYNELETRVLSKRAKAGVQSGTNALL-----VVK 352
; DB 573 DFGSDSG-----SDSNSDSGSDGSDTSDSGSDSAS 610

QY 353 HRDMNEKELEAQAARKAQLNHNPEPEEEETEKEACGSGDEEEOEKSGSSEKEGSEDE 412
; DB 611 DSDSASDSDSASDSDSASDSDSDNDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 670
; QY 413 HSGSESEFE--EGDRDEASDKSGSGEGESSEDEAARAARDKEEIFGSDADSDSDADSDDEDR 471
; DB 671 DS 729
; QY 472 GQAQGSNDSDSGSGGQGRSRSHSRASAPFFPSGSHSAQEDGSEAAASD\$-SEADSDS 530
; DB 730 SDS 788
; QY 531 D 531
; DB 789 D 789
; RESULT 13
; US-08-781-986A-5249
; Sequence 5249, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 5249:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-781-986A-5249
; Query Match 8.2%; Score 225.5; DB 4; Length 936;
; Best Local Similarity 21.4%; Pred. No. 2e-10;
; Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;

Qy 196 KPRVTPVEMVPPDFPKMWINPCAQVIFDSDPAPKOTSGAAALEMM-----SQMIRGM- 249
Db 477 NPNQYKVEFNT--PDDQITTPYIYVAVGHIDP---NSKGLALRSLTYGYSNIIWRSMS 531
Qy 250 MDEGNQFVAYFL-----PVEETLKKRKRDOEBENDYAPDDVVDYKIAREYNWNVKKA 303
Db 532 WONE---VAFNNGSGGDGDKPVPEQDPGEIEPIED-----SDS 572
Qy 304 SKGYEENYFFIFREGDGYVYNELETRVRLSKRAKAGVQSGTNALL-----VVK 352
Db 573 DPGSDSG-----SDSNDSGSDSGSDSTSDSGSDASDSAS 610
Qy 353 HRDMNKELEAQEARKAQLNHEPEEBEEMETEKEAGSGDEEKEKSGSSEKSGEDE 412
Db 611 DSDASDSASDSASDSASDSNDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 670
Qy 413 HSGSESERE-EGDRDEASDKSGSGEDESSEDEAARAARDKEEIFGSDADSDDDADSDDEDR 471
Db 671 DSD 729
Qy 472 GQAQGGSDNDSGSGGQGRSHRSASAPFFGSGSEHSAQEGSEAAASDS--SEADSDS 530
Db 730 SD 788
Qy 531 D 531
Db 789 D 789

RESULT 14
US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Edholm, Delidre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-5

Query Match 8.1%; Score 223.5; DB 4; Length 1315;
Best Local Similarity 21.7%; Pred. No. 4.9e-10;
Matches 121; Conservative 85; Mismatches 225; Indels 127; Gaps 23;

Qy 27 SGVYCRVKYCNLSPIPPDPKFTIYFPDQNRFOYKATSLKQKHDLLEPDLGVITDL 86
Db 772 TGVV-----NGADMTLDSGF--YKTPKNLGNVWEDTNKQKQSTEGISGVTVTL 823
Qy 87 INPD-----TYRID-----PNVLLDPADEKLEEEIQAP 115
Db 824 KNEGEVLQTTKIDCKGYQFTGLENGTYKVEPPTSGYPTQVGSCTDEG-IDSNGTST 882
Qy 116 TSSKRSQOAHKV-----VPMWRKTEYISTENRYGISENEKPEVKIGSVKQQTETEEIYK 170
Db 883 TGVIKDKNDTIDEGFYKPTYNLGDYVWEDTNKNGVQDKDEKIGISGVTV-----TLK 934
Qy 171 DRDSQITAIKTEFDAQKSISQ-----HYSKPRVTPVEMVPPDFPKMWINPCAQVIFDSD 226

Db 935 DENDKVLKVTITDENGKYQFTDLNNGTYKVBEPTSGYTP-----SVTSGN 981
Qy 227 PAPKOTSGAAALEMMSQAMIRGMDDEG-----NOFVAYFLPVEETLKKRKRDOEB 277
Db 982 DTEKDSNGLTTTGVTKDA--DNMTLDSGFYKTPKYSGLGDYVWY-----DSNKGKQDSTE 1034
Qy 278 EMDYAPDDVVDYKIAREYNWNVKNK--ASKGYEENYFFIFREGDGYVYNELETRVRLSKR 335
Db 1035 K-----GKIDVKVTI---LNEKGEVIGTKTDENKCYCFDNLDSKY-----KVIFEK- 1079
Qy 336 RAKAGV-QSGTNALLVVKHRDMNKELEAQEARKAQLNHEPEEBEEMETEKEAGGS 394
Db 1080 --PAGLTGTGTNTTDDKADGGEVDVTITDHDHDFLDNGYVEETSD-----S 1126
Qy 395 DDEQKSGSSEKEGSEDEHSGSESEREEDRDEASDKSGSGEDESSEDEAARAARDKEIF 454
Db 1127 DSD 1184
Qy 455 GSDADSEDDADDDRGOAQGGSDNDSGSGGQGRSHRSASAPFFGSGSEHSAQED 514
Db 1185 DSD 1237
Qy 515 GSEAAASDS--SEADSDSD 531
Db 1238 SD 1255

RESULT 15
US-09-949-016-10366
; Sequence 10366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10366
; LENGTH: 1259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10366

Query Match 8.0%; Score 220.5; DB 4; Length 1259;
Best Local Similarity 23.1%; Pred. No. 8.3e-10;
Matches 93; Conservative 57; Mismatches 160; Indels 93; Gaps 16;

Qy 171 DRDSQITAIKTEFDAQKSISQHYSKPRVTPVEMVPPDFPKMWINPCAQVIFDSDPAPK 230
Db 297 DHDSST-----GQNSDSKEYYDPEGK-----DPHNEV--DGDKTSK 331
Qy 231 DTSGAAALEMMSQAMIRGMDDEGNOFVAYFLPVBETLK-----KRQDEE 278
Db 332 SEENSA-----GIPDNGSQ-----RIEDTKLNHRESKRVENRTKSETHA 374
Qy 279 MDYAPDDVVDYKIAREYNWNVKNKASKGYE-----ENYFFIFREGDGYVYNE--LETRV 330
Db 375 VGSQDKGFEIKGPSSGSRNITKEVGKNEGKDGQHGMLGKGNVKTQGEVNVIEGPG 434
Qy 331 RLSKRAKAGVQSGTNALLVVKHRDMNKELEAQEARKAQLNHEPEEBE-----E 382
Db 435 QKSEPCNKVG-HSNTGS-----DSNSDGYSDYDFDKSKMQGDDPNSSDESGNDDANS 486
Qy 383 ENMETEKEAG-----GSDEQEKSGSSEKSGSEDEHSGSSEEREEDR-----DEASDK 431

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:48:56 ; Search time 174 Seconds
(without alignments)
1242.129 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAP1TQOARQEDGHRPNH.....QEDGSEAAASDSEADSDSD 531

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 1826554 seqs, 40705358 residues

Total number of hits satisfying chosen parameters: 1826554

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US11_PUBCOMB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764	100.0	531	9	US-09-840-787-9
2	2764	100.0	531	17	US-10-721-553-2
3	2658.5	96.2	553	18	US-10-450-763-50041
4	2464	89.1	473	10	US-09-986-480-410
5	2464	89.1	473	17	US-10-863-332-410
6	622	22.5	133	18	US-10-450-763-50040
7	452	16.4	86	15	US-10-424-599-213359
8	375.5	13.6	571	15	US-10-424-599-213359
9	337.5	12.2	644	16	US-10-437-963-116147
10	283	10.2	475	18	US-10-450-763-49771
11	238.5	8.6	286	16	US-10-425-115-202470

12	237.5	8.6	1633	15	US-10-282-122A-70437	Sequence 70437, A
13	234.5	8.5	1742	15	US-10-615-383-4	Sequence 4, Appli
14	234.5	8.5	1742	16	US-10-690-184-4	Sequence 4, Appli
15	234.5	8.5	1742	16	US-10-689-082-4	Sequence 4, Appli
16	234	8.5	953	15	US-10-282-122A-44457	Sequence 44457, A
17	233.5	8.4	930	16	US-10-744-672-3	Sequence 3, Appli
18	233.5	8.4	930	17	US-10-744-616-3	Sequence 3, Appli
19	233.5	8.4	947	17	US-10-470-048B-86	Sequence 86, Appli
20	233	8.4	932	9	US-09-815-242-5578	Sequence 5578, Ap
21	233	8.4	932	9	US-09-815-242-12438	Sequence 12438, A
22	230.5	8.3	1385	15	US-10-282-122A-44324	Sequence 44324, A
23	230	8.3	877	15	US-10-282-122A-70428	Sequence 70428, A
24	229	8.3	1920	15	US-10-282-122A-71413	Sequence 71413, A
25	228	8.2	428	16	US-10-451-467A-448	Sequence 448, App
26	226	8.2	913	17	US-10-470-048B-63	Sequence 63, Appli
27	226	8.2	918	16	US-10-744-672-1	Sequence 1, Appli
28	226	8.2	918	17	US-10-744-616-1	Sequence 1, Appli
29	225.5	8.2	933	17	US-10-470-048B-93	Sequence 93, Appli
30	225.5	8.2	936	8	US-08-781-986A-5249	Sequence 5249, Ap
31	225.5	8.2	936	15	US-10-329-624-5249	Sequence 5249, A
32	224.5	8.1	194	18	US-10-450-763-41624	Sequence 41624, A
33	224	8.1	265	13	US-10-073-256-78	Sequence 5, Appli
34	223.5	8.1	1315	16	US-10-744-672-5	Sequence 5, Appli
35	223.5	8.1	1315	17	US-10-744-616-5	Sequence 5, Appli
36	223.5	8.1	1315	17	US-10-470-048B-124	Sequence 124, App
37	221.5	8.0	1349	9	US-09-815-242-5898	Sequence 5898, Ap
38	221.5	8.0	1349	9	US-09-815-242-13137	Sequence 13137, A
39	220.5	8.0	1253	14	US-10-363-798-2	Sequence 2, Appli
40	220	8.0	717	17	US-10-470-048B-58	Sequence 58, Appli
41	219	7.9	1021	9	US-09-815-242-5471	Sequence 5471, Ap
42	219	7.9	1021	9	US-09-815-242-12544	Sequence 12544, A
43	218	7.9	406	16	US-10-451-467A-262	Sequence 262, App
44	218	7.9	1166	16	US-10-744-672-7	Sequence 7, Appli
45	218	7.9	1166	17	US-10-744-616-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-840-787-9
; Sequence 9, Application US/09840787
; Patent No. US20020058264A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: P1TUNOR01
; CLONE: 98974
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 9 :
US-09-840-787-9

Query Match      100.0%; Score 2764; DB 9; Length 531;
Best Local Similarity 100.0%; Pred. No. 9.8e-167;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPTIQTAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Db 1 MAPTIQTAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Qy 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
Qy 181 KTFEDAQKSIQSHYKPRVTPVEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQSHYKPRVTPVEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Db 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQARKAOLENHEPEEEEEETEKEAGGSDDEQEKGSSEKESGSESESER 420
Db 361 LEAQARKAOLENHEPEEEEEETEKEAGGSDDEQEKGSSEKESGSESESER 420
Qy 421 EBGDRDEASDKSGSEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGGSDN 480
Db 421 EBGDRDEASDKSGSEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGGSDN 480
Qy 481 DSDSGSGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
Db 481 DSDSGSGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 2
US-10-721-553-2
; Sequence 2, Application US/10721553
; Publication No. US20050032079A1
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; APPLICANT: University of Nebraska Board of Regents
; TITLE OF INVENTION: Novel Gene That is Amplified and
; TITLE OF INVENTION: Overexpressed in Cancer and Methods of Use Thereof
; FILE REFERENCE: UNMC63124
; CURRENT APPLICATION NUMBER: US/10/721,553

```

```

;
; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/647,143
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US99/06633
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/079,649
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
;
; US-10-721-553-2

Query Match      100.0%; Score 2764; DB 17; Length 531;
Best Local Similarity 100.0%; Pred. No. 9.8e-167;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPTIQTAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Db 1 MAPTIQTAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Qy 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
Qy 181 KTFEDAQKSIQSHYKPRVTPVEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQSHYKPRVTPVEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Db 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQARKAOLENHEPEEEEEETEKEAGGSDDEQEKGSSEKESGSESESER 420
Db 361 LEAQARKAOLENHEPEEEEEETEKEAGGSDDEQEKGSSEKESGSESESER 420
Qy 421 EBGDRDEASDKSGSEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGGSDN 480
Db 421 EBGDRDEASDKSGSEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGGSDN 480
Qy 481 DSDSGSGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
Db 481 DSDSGSGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 3
US-10-450-763-50041
; Sequence 50041, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736

```

SOFTWARE: Custom
SEQ ID NO 50041
LENGTH: 553
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (377)..(428)
OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
OTHER INFORMATION: accession number BL00412D, p-values=9.633e-09, raw score of 16.54
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(553)
OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50041

Query Match 96.2%; Score 2658.5; DB 18; Length 553;
Best Local Similarity 95.0%; Pred. No. 4.9e-160;
Matches 515; Conservative 6; Mismatches 10; Indels 11; Gaps 1;

Qy 1 MAPTIQAOQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Db 12 MAPTIQAOQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 71
Qy 61 YKATSLKQKHDKHLLTEPDGLVTIDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 72 YKATSLKQKHDKHLLTEPDGLVTIDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 131
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQOFTTEEYIKRDSQITAE 180
Db 132 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQOFTTEEYIKRDSQITAE 191
Qy 181 KTFEDAQKS-----ISQHSKPRVTPVEVMPVFPDFKWINPACQVIFDSDPAP 229
Db 192 KTFEDAQKSVIEGLGWGEARISQHSKPRVTPVEVMPVFPDFKWINPACQVIFDSDPAP 251
Qy 230 KDTSGAAALEMMSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKDOEEMDYAPDDVYDY 289
Db 252 KDTSGAAALEMMSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKDOEEMDYAPDDVYDY 311
Qy 290 KIAREYNWNVKSKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALL 349
Db 312 KIAREYNWNVKSKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALL 371
Qy 350 VVKHRDMNEKELEQAARKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKGS 409
Db 372 VVKHRDMNEKELEQAARKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKGS 431
Qy 410 EDEHSGSESEREEDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDE 469
Db 432 EDEHSGSESEREEDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDE 491
Qy 470 DRGOAGGSDNDSDGRNGGGGQTRSHRSASRSPFSGSEHSAQEGSEAAASDSSEADSD 529
Db 492 DRGOAGGSDNDSDGRNGGGGQTRSHRSASRSPFSGSEHSAQEGSEAAASDSSEADSD 551

Qy 530 SD 531
Db 552 SD 553

RESULT 4
US-09-986-480-410
Sequence 410, Application US/09986480
Publication No. US2003002199A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 143 Human Secreted Proteins
FILE REFERENCE: PS500P1
CURRENT APPLICATION NUMBER: US/09/986,480
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: PCT/US00/12788
OTHER INFORMATION: Xaa = X or * as defined in Table 2

PRIOR APPLICATION NUMBER: US 60/134,068
PRIOR FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 456
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 410
LENGTH: 473
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (405)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-986-480-410

Query Match 89.1%; Score 2464; DB 10; Length 473;
Best Local Similarity 99.8%; Pred. No. 8.3e-148;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAPTIQAOQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Db 1 MAPTIQAOQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Qy 61 YKATSLKQKHDKHLLTEPDGLVTIDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 61 YKATSLKQKHDKHLLTEPDGLVTIDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQOFTTEEYIKRDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQOFTTEEYIKRDSQITAE 180
Qy 181 KTFEDAQKSISQHSKPRVTPVEVMPVFPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSISQHSKPRVTPVEVMPVFPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKDOEEMDYAPDDVYDYKIAREYNWNVK 300
Db 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKDOEEMDYAPDDVYDYKIAREYNWNVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQEAARKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKGSSEHSGSESER 420
Db 361 LEAQEAARKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKGSSEHSGSESER 420
Qy 421 EBGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDEDDRSQ 473
Db 421 EBGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDEDDRSQ 473

RESULT 5
US-10-863-332-410
Sequence 410, Application US/10863332
Publication No. US20050064458A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 143 Human Secreted Proteins
FILE REFERENCE: PS500P1
CURRENT APPLICATION NUMBER: US/10/863,332
CURRENT FILING DATE: 2004-06-09
PRIOR APPLICATION NUMBER: US/09/986,480
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: PCT/US00/12788
PRIOR FILING DATE: 2000-05-11
PRIOR APPLICATION NUMBER: US 60/134,068
PRIOR FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 456
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 410
LENGTH: 473
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (405)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-863-332-410

Query Match 89.1%; Score 2464; DB 17; Length 473;
Best Local Similarity 99.8%; Pred. No. 8.3e-148;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAPTIQTAQREDGHRNPSHTLPSRGVCRVKYCNLSLPDIPDPKFIYTPDQNRVQ 60
Db 1 MAPTIQTAQREDGHRNPSHTLPSRGVCRVKYCNLSLPDIPDPKFIYTPDQNRVQ 60

Qy 61 YKATSLKQKHDLITPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKR 120
Db 61 YKATSLKQKHDLITPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKR 120

Qy 121 SQQAKVVPVNRKTEYISTENRYGIGNEKPEVKIGSVKQOFTTEBEIYKDRDSQITAE 180
Db 121 SQQAKVVPVNRKTEYISTENRYGIGNEKPEVKIGSVKQOFTTEBEIYKDRDSQITAE 180

Qy 181 KTFEDAQKSIQHSKRVTPVEVMPVFPDFKMWINPCAOVIFDSDPAPKDTSGAALEM 240
Db 181 KTFEDAQKSIQHSKRVTPVEVMPVFPDFKMWINPCAOVIFDSDPAPKDTSGAALEM 240

Qy 241 MSQAMIRGMDDEGNQFVAVLPVEETLKRRQRQDEEMDYAPDDVDYDKIAREYNWVK 300
Db 241 MSQAMIRGMDDEGNQFVAVLPVEETLKRRQRQDEEMDYAPDDVDYDKIAREYNWVK 300

Qy 301 NKASKGYEENYFFIFRGDGVYVNELETRVLSKRAKAGVQSGTGNALLVVKHRDMNKE 360
Db 301 NKASKGYEENYFFIFRGDGVYVNELETRVLSKRAKAGVQSGTGNALLVVKHRDMNKE 360

Qy 361 LEAQEARKAQLENHEPEEEEEETBEKAGGSDERQEGSGSSEKSGSEDEHSGSESER 420
Db 361 LEAQEARKAQLENHEPEEEEEETBEKAGGSDERQEGSGSSEKSGSEDEHSGSESER 420

Qy 421 EBGDRDASDKSGSEDESSDEARAAKDEEIFGSDADSDDDSDDRGQ 473
Db 421 EBGDRDASDKSGSEDESSDEARAAKDEEIFGSDADSDDDSDDRGQ 473

RESULT 6
US-10-450-763-50040
; Sequence 50040, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60/736
; SOFTWARE: Custom
; SEQ ID NO 50040
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-50040

Query Match 22.5%; Score 622; DB 18; Length 133;
Best Local Similarity 64.4%; Pred. No. 7e-32;
Matches 130; Conservative 0; Mismatches 2; Indels 70; Gaps 1;

Qy 273 RDQEEEMDYAPDDVDYDKIAREYNWVKKASKGYEENYFFIFREGDGVYVNELETRVRL 332

Db 1 ROOEEEMDYAPDDVDYDKIAREYNWVKKASKGY----- 35
Qy 333 SKRRAKAGVQSGTGNALLVVKHRDMNEKELEAQAQKAELENHEPEEEEEETBEKEAG 332
Db 36 -----EEEEEETBEKEAG 50

Qy 393 GSDEOEKSGSSEKSGSEDEHSGSESEREGDRDEASDKSGSEDESSDEARAAKDEE 452
Db 51 GSVEOEKSGSSEKSGSEDEHSGSESEREGDRDEASDKSGSEDESSDEARAAKDEE 110

Qy 453 IFGSDADSDDDSDDDDRGQA 474
Db 111 IFGSDADSDDDSDDDDRGQA 132

RESULT 7
US-10-424-599-223174
; Sequence 223174, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223174
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43557C.1.pap
US-10-424-599-223174

Query Match 16.4%; Score 452; DB 15; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.4e-21;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 LEKQKHDLITPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKRQQHA 125
Db 1 LEKQKHDLITPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKRQQHA 60

Qy 126 KVPVNRKTEYISTENRYGIGNEKP 151
Db 61 KVPVNRKTEYISTENRYGIGNEKP 86

RESULT 8
US-10-424-599-213359
; Sequence 213359, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 213359
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34688C.1.pap
US-10-424-599-213359

Query Match 13.6%; Score 375.5; DB 15; Length 571;
Best Local Similarity 27.2%; Pred. No. 1.6e-15;
Matches 128; Conservative 71; Mismatches 170; Indels 101; Gaps 19;

Qy 14 GHRPNHSTLP-----ER-----SGVVCVKYCNLSLPDPFPKFTYFPDQNRFPVQY 61
Db 150 GSRMGERSSTPLGAAERVENRLKPTTFLCKLKFNLDPDPSAOKPLMASKKDQVAKY 209
Qy 62 KATSLKQKHLDLITPDLGVTDILINPDTYRIDPNVL--LDPADKLLLEE-----IQAA 114
Db 210 TITSLKMYKPKLFWPDLGIPDLDDLVLVYN-PSVAPPLAPEDKELLRDEAVTPIKK 268
Qy 115 PTSSKRSQOQAKVPMWKRTEYISTEFNRYGISNEKPEVKIGSVKQOFTFEEI-----168
Db 269 DGIKRKERTDKGVAWLKTKVQVISP-----LSME-----STKQSLTEKQAKELREM 314
Qy 169 -----YKDRSQITAEKTPEDAKSISQHVSKPRVTPVEVMPFDPKMWNPCA 219
Db 315 KGGRILOLNSRERQIREIASFE-AAKSDPVHATNKDLYPVEVMPPLLPDPDRYDQFV 373
Qy 220 QVIFSDP-----APKDTSGAAALE---MMSQAMIRGMWDEEGNQFVAYFLPVEETLK 269
Db 374 VAFNATATASEMAKVKDKSVDFESKAVKSVATGSDPANPEKFLAYWVPAGELS 433
Qy 270 KRKRDOEBEMDYPDDVYDYKIAREYNNVKNKASKGYEENYFFIFREGDGVYNELETR 329
Db 434 KOIYDENEVSYS-----WIREYHMDVRGDDADD-PATFLVAFDSEARYL-PLPTK 483
Qy 330 VRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEMEETEEK 389
Db 484 LVLRKKRAKEG-RSG-----DEVEQCFVPARVTVRRSSVAAIERK 523
Qy 390 RAGGSDEQEKGSSSEK-----EGSEDEHSGS---ESEREGDRDEASD 430
Db 524 DSG--VYTTSSKNSKRGKLEMDGLEDQHRGPHQDNVYQSSGAEDYMSD 571

RESULT 9
US-10-437-963-116147
; Sequence 116147, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204566
; SEQ ID NO 116147
; LENGTH: 644
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19676C.1.pap
US-10-437-963-116147

Query Match 12.2%; Score 337.5; DB 16; Length 644;
Best Local Similarity 25.3%; Pred. No. 4.7e-13;
Matches 109; Conservative 82; Mismatches 156; Indels 83; Gaps 17;

Qy 17 PMSHR---TLPRSGLVCEVKYCNLSLPDPFPKFTYFPDQNRFPVQYKATSLKQKHHD 73
Db 228 PNAERFENRLKPTTFLCKHKNLDPDPSOLKWLPLNKOKORYTKRITSLEKNYIPK 287

Qy 74 LLTEPDLGVTDILINPDTYRIDPNVL-----LD 101
Db 288 MIVPEDLGIFLDLDMVYKIELSQIYVFLAAILRIRFFLAYLTYAOTIFSTPPVQPPMA 347
Qy 102 PADEKLL-BEEOIAPTSS---KRSQOQAKVPMWKRTEYI---STBFNRYEI---SNE 149
Db 348 PEDEELLARDDDEVLPVKDGIRKERTDKGMSLVKTVIISPLSTDAAKMSITEKQAKE 407
Qy 150 KPEVKIGSVKQOFTFEEIYKDRDSQITAEKTPEDAKSISQHVSKPRVTPVEVMPVFP 209
Db 408 RRESREG---RNTFLEN--INDREKQIKAIEDSFR-AKSRPVHQTQKRGMEAEWVLPPLP 461
Qy 210 DEFKWINPCAQVLFSDPAKDTSGAAALE-----MMSQAMIRGMWDEEGNQFV 258
Db 462 DFORYDDQFVWVNFDPDPT-ADSEQYNNKLSERSEDECSRAVMKSLVNGSDPAKQKFL 520
Qy 259 AYFLPVEETELKGRDOEBEMDYPDDVYDYKIAREYNNVKNKASKGYEENYFFIFREG 318
Db 521 AYWVPSHELKDLDDDETIDQYS-----WLREYHMEVVRGD-DKDDPTTYLVTF-DD 570
Qy 319 DGYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLN---HE 375
Db 571 DGAKYLPPLTKLVQKKKAKEG-RSGDE-----IEHFPVPSRITENLKRQSSVDDDLVDH 625
Qy 376 PEEEEEEME 385
Db 626 PKHSRVEDMD 635

RESULT 10
US-10-450-763-49771
; Sequence 49771, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 05/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49771
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (391)..(429)
; OTHER INFORMATION: ELEMENT TRANSPOSABLE INSERTION PROTEIN TRANSPOSITION DNA
; OTHER INFORMATION: domain identified by eMATRIX, accession number PD02455A, p-value:
; OTHER INFORMATION: 1.450e-25, raw score of 25.65
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (68)..(113)
; OTHER INFORMATION: Immunoglobulin domain identified by Pfam, accession name ig,
; OTHER INFORMATION: E-value=0.099, Pfam score of 10.6
US-10-450-763-49771

Query Match 10.2%; Score 283; DB 18; Length 475;
Best Local Similarity 87.7%; Pred. No. 9.2e-10;
Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 328 TRVLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEMEETE 387
Db 3 SRVLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEMEETEIRQP 62
Qy 388 EKEAG 392

Qy 469 EDRGQAQGGSDNDSDSGNGGQGRSRSHRSASPPFPGSGSEHSAQEDGSEAAADS-SEAD 527
Db 1230 DSDAD 1286
Qy 528 SDSA 531
Db 1287 SDSA 1290

Qy 469 EDRGQAQGGSDNDSDSGNGGQGRSRSHRSASPPFPGSGSEHSAQEDGSEAAADS-SEAD 527
Db 1230 DSDAD 1286
Qy 528 SDSA 531
Db 1287 SDSA 1290

RESULT 15
US-10-689-082-4
; Sequence 4, Application US/10689082
; Publication No. US20040142348A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS
; FILE REFERENCE: P06335US04/BAS
; CURRENT APPLICATION NUMBER: US/10/689,082
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: 09/386,962
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-689-082-4

Query Match 8.5%; Score 234.5; DB 16; Length 1742;
Best Local Similarity 21.0%; Pred. No. 5.1e-06;
Matches 127; Conservative 88; Mismatches 240; Indels 149; Gaps 24;
Qy 48 PITVPPONRFVQKATSLKQKHDLTPDGLVTTDLINPOTYRIDP-----NVLDP 102
Db 716 YVTLKDSNNRELQVTTDQSGHQFNQNGT--YTVEFAIPDNYPSPANNSTNDAIDS 773
Qy 103 ADEKLEEEIQAPTSKRSQKAKV-----VPMRTEYISTEFNRYGIGSNEKPEVKIG 156
Db 774 DGERDGRKVVAKGTINNADNMVTGTGFLTPKYNVGVYVWEDTNKGIQDDNEKIGS 833
Qy 157 VSV-----KQOFT--EEIY-----KDRSQ 175
Db 834 VKVTLKKNKNGDTIGITTTSDNGKYEFTGLENGDYTIEFETPEGYTPKQNSGDEGKDSN 893
Qy 176 ITALEKTFEDA-QKSIQHYKPRVTPVMPVPPDFPMINPCAVIFSDP-----227
Db 894 GTKTVTVKQADNKTIDSGFYKPTYN-----LGDY-VWEDTNKGIQDDSEKIGSGVK 945
Qy 228 -APKDTSGAA---ALEMMSQAMIRGMADDEGNQFVAYFLPVEBTLKRRKQDEEMDY - 281
Db 946 VTLKDKNGNALGTTTDTASGHYQKGL--ENGSTVEFETPSGYTPKANSQQDITVDSN 1003
Qy 282 -----APDDVYD---YKIAR---BYNMNVKXK-----ASKGY-----307
Db 1004 GIITTTGIINGADNLITDSGFYKTPKYSVGVYVWEDTNKGIQDDNEKIGSVKVTLKDEK 1063
Qy 308 -----BENYFFIFREGD-GVYNELETRVRLSKRAKAGVQSGCTNALLVVKHRDMN 357
Db 1064 GNIISTTTTIDENGKYPQDNLDSNGYIIHFKEKPGMTOTTANS-----NDD 1109
Qy 358 EKELEAQEAKA-----QLEHPEPEEEEEETEKEAGGSGDEQEKGSSSEKESGE 410
Db 1110 EKADADGEDVRVTIDHDDFIDNGYFDSDSDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDA 1169
Qy 411 DEHSGSEERE-ECDREASDK-SGSGEDSSSEDEAARAARDKEEIFGSDADSDSDSDSD 468
Db 1170 DS 1229

Search completed: September 26, 2005, 13:59:26
Job time : 176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:44:30 ; Search time 42 Seconds
(without alignments)
1216.455 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPIQQAQREDGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	645	23.3	425	2	T20261	hypothetical prote
2	335.5	12.1	547	2	C36828	unknown protein F1
3	280.5	10.1	457	2	T50233	probable DNA-direc
4	253	9.2	445	2	S44541	hypothetical prote
5	245.5	8.9	784	2	PN0009	neurofilament trip
6	241	8.7	792	2	T42963	hypothetical prote
7	236	8.5	590	2	A40437	glutamic acid-rich
8	234	8.5	953	2	C89824	hypothetical prote
9	230.5	8.3	1385	2	D89824	hypothetical prote
10	230	8.3	877	2	F90070	Clumping factor B
11	225.5	8.2	933	2	A51539	fibrinogen-binding
12	224	8.1	797	2	A36811	hypothetical prote
13	223.5	8.1	1315	2	T28679	fibrinogen-binding
14	221	8.0	1192	2	A1623	probable secreted
15	220	8.0	334	2	A54138	acidic repetitive
16	218	7.9	406	2	S38170	SRP40 protein - ye
17	218	7.9	1166	2	T28680	fibrinogen-binding
18	216	7.8	913	2	T52485	neurofilament prot
19	216	7.8	1110	2	I51116	NF-180 - sea lamp
20	214.5	7.8	1141	2	E89824	hypothetical prote
21	214	7.7	727	2	JC5113	ribosomal transcri
22	214	7.7	2364	2	A56577	microtubule-associ
23	213.5	7.7	678	2	A54514	glutamic acid-rich
24	212.5	7.7	1092	2	T30214	fibrinogen-binding
25	211.5	7.7	665	2	B71609	hypothetical prote
26	211	7.6	985	2	D89852	fibrinogen-binding
27	208.5	7.5	798	2	I50479	neurofilament medi
28	208	7.5	765	2	S22314	transcription fact
29	203.5	7.4	606	2	S70358	centromere protein

ALIGNMENTS

RESULT 1

T20261

hypothetical protein C55A6.9 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T20261

R;Kershaw, J.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19243

A;Accession: T20261

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-425 <WIL>

A;Cross-references: UNIPROT:P90783; EMBL:Z81051; PIDN:CAB02869.1; GSPDB:GN00023; CESP:C

A;Experimental source: clone C55A6

C;Genetics:

A;Gene: CESP:C55A6.9

A;Map position: 5

A;Introns: 14/2; 48/2; 90/3; 177/3; 381/1

Query Match 23.3%; Score 645; DB 2; Length 425;

Best Local Similarity 33.1%; Pred. No. 3.8e-26; Indels 34; Gaps 10;

Matches 146; Conservative 96; Mismatches 165;

Qy 24 PERSGVVCRVKYCNLSLPDIPDPKFIITYPF-DQNRVQYKATSLKQKHDLTEPLGV 82

Db 14 PRKVDPMKPRFTNTVPDVFDFDAKFWTCFVPLGRFVFQPAIYRDYKHAVICDDMGL 73

Qy 83 TIDLINPDYRIDP-NVLLDPADEKLEEBEIQAPTSSKRQOQAKVVPWVKTEYISTEF 141

Db 74 NVLDLIDLKDYDEPIETEIDEKDNILLDDGAAKLIAKRSQOQSKLVFPWVKTEYISTEF 133

Qy 142 NRGISNEKEVGLSVKQFTTEELVKORDSOITAEKTFEDAQKSIQSYKPRVTP 201

Db 134 NRGVTDADROETKGLYNLKNQOVEDMYRDQKQIDAINKTFEDVRKPKVKEHYKGVKA 193

Qy 202 VEYMPVFPDFKWINPCAQVIFSDPAKPDTSAGAALEMMQAMIRGMDEEGNQFVAYF 261

Db 194 VESFVFPDFDHWKHLFAHFQFDGDTITTFEEDERQQARESSVIKAMEFDDQKFAAVF 253

Qy 262 LPVEETLKKRKDOEEEMDYPDDVDYKTAERYNNWNKNAKSGYENYFIIFREGDV 321

Db 254 VPTIGCLTSPMDLLELRPFEDMKVEFLLSREYTFKMEHLPPR--DRDVFIMYHRNVVF 311

Qy 322 YNLETRVLSKRRKAGVQSGTNALLVVKHRDMKELEAQARKAQLNHE-PEEE 380

Db 312 QYNEVDCNVKMT-RKPKVALSRKSKULTYR-----NFSLEKQDMNKREALYIQPKTRK 366

Qy 381 EEME--TEEKEAGSGDEQEKSGSSSEKSGSEDEHSGSESESEDEGRDEADSKSGSEDE 438

Db 367 QEILEKIQEKKEGGD-----SSDQSSDSDDDKPKQSR-----SDSSSDV 406

A:Molecule type: DNA
A:Residues: 1-445 <HOL>
A:Cross-references: UNIPROT:P38351; EMBL:X76053; NID:G600025; PIDN:CAA53642.1; PID:G4291
R:Brandt, T.; Christiansen, C.; Holmstroem, K.; Kalliesoe, T.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S46157
A:Accession: S46161
A:Molecule type: DNA
A:Residues: 1-445 <BRA>
A:Cross-references: EMBL:Z36548; NID:G536721; PIDN:CAA85243.1; PID:G536722; MIPS:YBR279W
R:Shi, X.; Finkelstein, A.; Wolf, A.J.; Wade, P.A.; Burton, Z.F.; Jaehning, J.A.
Mol. Cell. Biol. 16, 669-676, 1996
A:Title: Paf1, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may
A:Reference number: JG6088; MUID:96140434; PMID:8552095
A:Accession: JG6088
A:Molecule type: DNA
A:Residues: 1-166,168-445 <SHL>
A:Experimental source: strain YJ3453
A:Accession: PC6031
A:Molecule type: DNA
A:Residues: 5-11;420-427 <SH>
C:Comment: This factor is a highly charged nuclear protein, and acts as a cofactor impor
C:Genetics: PAF1
A:Gene: SGD:PAF1
A:Cross-references: SGD:S0000483; MIPS:YBR279W
A:Map position: 2R
A:Note: this gene is located at the right arm of chromosome II
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YBR279W
C:Keywords: nucleus
F:25-49/Region: PEST sequence
F:119-141/Region: nuclear location signal

Query Match 9.2%; Score 253; DB 2; Length 445;
Best Local Similarity 22.1%; Pred. No. 4,4e-06;
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

Qy 23 LPERSGVCRVYCNLSLDPIDPPDKFITP-----FDQNRFOYKATSLKQHK 71
Db 1 MSKQEXIAPIKYQNSLPVLPKLLVYSPETNADSSQLNSLYIKTNVTLIQ-- 58
Qy 72 HDLLEPDLGTTIDL-----NPDYTRIDPNVLLDPADEKLLEETQATSSKRS 121
Db 59 -----DEDLGHPVLDLKPFGGLNKLDSKLLYGFD-NVKLKDDRIILLRD----PRIDRLT 108
Qy 122 QOHAKVPMWPKTEIISTEENRYGISNEKPKVIGSVKQFTEEEIYKDRDSQ----- 175
Db 109 KIDISKVTLRTETFSNTIAHNTSLKRRRL-----DDGSDDENLDV 154
Qy 176 ---ITAIEKTFEDAKSISQHSYKPRVTPVPMVFPDPFKMWINPCAQVIFDSDPAKOT 232
Db 155 NHIISRVEGTFNKTK--WQHPVKGVKGVKWDLLPD-----TASMDQVYE-----ILKF 203
Qy 233 SGAAALEMMSQAMIR-GM---MDEGNQFVAYFLPVEETLKKRKDDQEEWDYAPDDVYD 288
Db 204 MGSASLDTKKKSLNTGIFRPVELEDEEWISMYATDHKDSAILNELEKGMDEMDDDSHE 263
Qy 289 ---YKIAEYNNVNVKASKGYEENYFFIFREGDGV-YNBELETRVLSKRRAKAGVQSG 344
Db 264 GKIIYKFKIRIDYDMQVAKPWTE-LATRLNDKOGIAYIKPLRSKIELRRRVNDIYKP- 321
Qy 345 TNALLVVKH-----RDMNEKELEAQAKAQLN-----HEPEEEEEEMETEET 389
Db 322 ---LVKEHDIDQLNVLRLNFTSTKEANIRDKLRFKDFDINFATVDEEDEDQEDVVK 377
Qy 390 EAGSDEFEQKSSSEKSGSDHSGSREBERGRDRBASDKSGSGEDESSEDEARAARD 449
Db 378 ESEG--DSKTEGSEGEENKDEEIKQEKENEQ-----DEENKQDENRAADT 422
Qy 450 KEETFGSDADSEDDADSDDEDRGQAQ 475
Db 423 PET---SDAVHTEQKPEEEKETLOEE 445

RESULT 5
PN0009
neurofilament triplet M protein - Pacific electric ray (fragment)
C:Species: Torpedo californica (Pacific electric ray)
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 09-Jul-2004
C:Accession: PN0009
R:Linial, M.; Scheller, R.H.
J. Neurochem. 54, 762-770, 1990
A:Title: A unique neurofilament from Torpedo electric lobe: sequence, expression, and lc
A:Reference number: PN0009; MUID:90155300; PMID:2106008
A:Accession: PN0009
A:Molecule type: mRNA
A:Residues: 1-784 <LIN>
A:Cross-references: UNIPROT:Q7LZ90
C:Comment: Neurofilaments are a subgroup of intermediate filaments which are expressed e
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; cytoskeleton; intermediate filament; nerve; phosphoprotein; tar
F:1-52/Region: serine-rich
F:53-84/Region: coil Ia
F:98-194/Region: coil Ib
F:217-367/Region: coil II
F:400-597/Region: glutamic acid-rich
F:598-674/Region: 6-residue repeats
F:675-784/Domain: carboxyl-terminal #status predicted <CTD>
F:616,622,628,634,640,646,652,658,670/Binding site: phosphate (Ser) (covalent) #status F

Query Match 8.9%; Score 245.5; DB 2; Length 784;
Best Local Similarity 20.5%; Pred. No. 2.1e-05;
Matches 122; Conservative 102; Mismatches 229; Indels 143; Gaps 23;

Qy 47 KPTYPFDQNRFOYKATSLKQ---HKHLLTEPDLG-----VTIDLINPDYTR 93
Db 62 RFAGY-IDKVHYLEQQNKLEAETQAHKQKQVSHGQLGDVYDQIRLRSIRQVNOEKAQ 120
Qy 94 ID-PNVLLD-----PADEKLLEEE-----IQAPTSSKSSQOHAKV 128
Db 121 IQDLSVHLDDFORVGFADBEALRDEPEATIRLVKKEDESVIQADGGEKPAQSLQDE 180
Qy 129 PMWRKTEYISTEENRYGISNEKPKVIGSVKQFTEEEIYKDRDSQITAIHKTFFDAQK 188
Db 181 AFLR-----NNHEEV-ADLFAQIQATQVTVKEK-KDFLKTDTISALKEIRS 224
Qy 189 SISOHSYKPRVTPVPMVFPDPFKM-----INPCAQVIFDSDPAKQDTS GAALAEEM 241
Db 225 QLECHSAKNMQADE-----WPKCRYDKLNEAAEMNKDAIRAAREEIGYRRQLQ 274
Qy 242 SQAM-----IRGM-----MDEGNQFVAYFLPVEETLKKRKDOEEM-----DYA 282
Db 275 SKSIELESVSTKESLERQLTDIEDRHADVANYQETVQQLLENELRGTKWENARHLREY- 333
Qy 283 PDDVYDYKIAEYNNVNVKASKGYEENYFFIFREGDGVYNN-----ELETRVLSKRR 337
Db 334 -QDLLNVKMALDIEIAAYRKLLGEESRYTF-SGTGPSIPYRSPSRPRLPAKVHKTKEVP 391
Qy 338 KAGVQSGTNALLVVKHRDMNEK-----ELEAQAKAQLNHEPEEEEEEMETE----- 388
Db 392 KVKYQHKFVEEIIETETKVQDEKAEAGMDIDLAEEVGGATMESPEDKEAEKVEEIVAT 451
Qy 389 -----KEAGSDEFEQKSSSEKSGSDHSGSREBERGRDRBASDKSGSGEDESSED 442
Db 452 VKGAVQAPPEGAEESESEAKKEEDEGVVEEKEKE--EADDEEKEEKEDEGEADAEAGG 510
Qy 443 EARAARDKEEIF-----GSDADSEDDADSDDEDRGQAQGGSDNDSSGNGGQQR- 492
Db 511 ESRVVEKVIKVKQSKAHPGKDEVKKEKEKEEKEEKEEKEEKEEKEEKEEKEEKEE 570
Qy 493 -----SRSHSRASGFPFSGSHSAQED-----GSEAAASDSSSEADSDS 530
Db 571 SKGKVEELTVKTEKATEDKVPREKPKQEKQDIEEKKEAKSKDEAKSKDEAK 626

RESULT 6
T42963

hypothetical protein 48 - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42963
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: 222274
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-792 <ALB>
A:Cross-references: UNIPROT:Q9YTL7; EMBL:AF083424; PIDN:AAC95573.1
A:Experimental source: strain 73

Query Match 8.7%; Score 241; DB 2; Length 792;
Best Local Similarity 23.2%; Pred. No. 3.5e-05;
Matches 94; Conservative 69; Mismatches 173; Indels 108; Gaps 14;

QY 169 YKRDSDQITAIKPFDAQKSI-----SQHYSKPRVTPVEMVFP 209
DB 213 YQYMSDLIAEELALQSLSICGSTYPSYSKILELLTANNSKEHIRQKVNVTD----- 266

QY 210 DFKWINPCA-QVTFDSDPAPKDTSGAAALEMWSQAMIRG-----MDSEGNO 256
DB 267 ----FIKPSLHQMFRTDCKGPRQKTKTLMISILGS-----RGIGDLFTQDVLKPPSSDAK 319

QY 257 FVAYFLP-----VEETLKKRKRDQEEEMDYAPDDVYVYKIAREYNMKNKASK----- 305
DB 320 FMAVSQDPNFKEKEVEFSMTGGKTSDDVT--APRKVGKSLARKYLENLKDKRKNY 377

QY 306 -GYEYNTFFIFREDG-----VYVNELE-----TRVLSKRRKAGVQSGTNAL-- 348
DB 378 SGRNKKY-----KDGANDKDKSIDKNESEGGHSEINREKNRKRKPNKPNVGDKEVGE 432

QY 349 -----LVVGHDRMNEKELEAQAQKLEHNEPEEEEEEMETEEK 399
DB 433 EKSVMSEGGKKEKSEERAEKDEENKKGGEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 492

QY 390 EAGSGDEQKSGSE--KEGSEDEHSGSESEEGDRDEASDKSGSGESESDEARAA 447
DB 493 EEEEEDE 552

QY 448 RKKEIFGSDADSDADSDRQAGGSDNDSDSGSGGQGRSHRSASPPFSGS 507
DB 553 EEEEEDE 605

QY 508 EHSQAQDGEAAASDSSEADSDSD 531
DB 606 EXEDEDDE 629

RESULT 7
A40437
Glutamic acid-rich protein, retinal - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 09-Jul-2004
C:Accession: A40437
R:Sugimoto, Y.; Yatsunami, K.; Tsujimoto, M.; Khorana, H.G.; Ichikawa, A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3116-3119, 1991
A:Title: The amino acid sequence of a glutamic acid-rich protein from bovine retina as deduced from complementary DNA
A:Reference number: A40437; MUID:91195303; PMID:2014230
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-590 <SUG>
A:Cross-references: UNIPROT:Q28181; GB:M61185; NID:g163077; PIDN:AAA30536.1; PID:g163078

Query Match 9.5%; Score 236; DB 2; Length 590;
Best Local Similarity 22.4%; Pred. No. 4.5e-05;
Matches 133; Conservative 80; Mismatches 221; Indels 160; Gaps 28;

QY 21 RILPERSGVVCRVKY--CNSLDPIDFPDPKFIYPPDQNRFPVQYKATSLKCHKHDLITTEP 78
DB 7 RVLPPGPTQKTKQEEEGTEPELEPKPTAPES-----TELEVSPLPE-----EP 55

QY 79 DLGVTIDLNPDIYRIDPNVLLDPADEKLEEBIQAPTS---SKRSQQAQVVPVPMRK-- 133
DB 56 CVGKEVAATLGPQGTQETALTPT-----SLQAQVSAPEAHSSPRGWLTWLRKGV 108

QY 134 -----TEYISTEFNRYGISNEKPEVKGVSVKQQTTEBEEIYKDRDSQTAIE----- 180
DB 109 EKVVPQPAHSSRPSONIAAGLESPPDQQAQILGCGTGG--SDEPSEPSRAEDPPGPGW 166

QY 181 --KTEE-DAQKISQHYSKPRVT-----PVEVMPVP----- 208
DB 167 LLRWFEQNLKRLPQ---PPKISEGWRDEPTDAALGPEPPGPALETKPMLQAQESPLPA 223

QY 209 -----PDFKWINPCAQVIFDSDPAPKDTSGAAA-----LEM-MSQAMIRGMMDEGNQF 257
DB 224 PGPPPEEPIPEPQPTIOASSLPPQDSARLWMLHLREWALPOPVIRKGGEGESD- 282

QY 258 VAYFLPVE---ETLKKRKRDQEEEMDYAPDDVYVYKIAREY--NMVKNKASKGYEENYF 312
DB 283 ---APVTCDVQTSILPGEQEE-----SHLLEEVDPHW----- 313

QY 313 FIFREGDGVYVNELETRVLSKRRKAGVQSGTNALLVVKH-----RDMNEKELEAQA 366
DB 314 -----BEDEHOGESTSPRTSE-AAAPADEKGE---VVEQTPRELPRLOEKEDEEEK 364

QY 367 RKAQLENHPEPEEEEEEMETEKEAGGSDDEQKSGSSSEKGESE-DEHSGSSESEREGDR 425
DB 365 EDGEEEEEGREKEEEEGEKEEEE-GREKEEEGKEKEEGREKEEEEGEKEDEGRE 423

QY 426 DEASDKSGSGESESSEDEARAARDEKKEIFGSDADSD-----DEDRG 472
DB 424 KEEEEGRGKEEKEEGEKEEGEKEEGREVEEGREDEEEDHSHVLLDSYLVQSEEDOS 483

QY 473 QAQGGSDNDSDSGSGGQGRSHRSASPPFSGSHSAQEDGSE-AAASDSSE 525
DB 484 E-----ESETQDQSEVGGAGTQGEVGGAAQL---SESETQDQSEVGGAGDQSE 529

RESULT 8
C89824
hypothetical protein edrc [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C89824
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C89824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-933 <KUR>
A:Cross-references: UNIPROT:Q99W48; GB:BA000018; PID:g13700453; PIDN:BA841750.1; GSPDB:
A:Experimental source: strain N315
C:Genetics:
A:Gene: edrc

Query Match 8.5%; Score 234; DB 2; Length 933;
Best Local Similarity 23.0%; Pred. No. 9.9e-05;
Matches 131; Conservative 81; Mismatches 216; Indels 142; Gaps 28;

QY 35 YCNSLPDIPFDP---KFIYTFP-DQNRFPVQYKATSLK-----QHKHDLITTEPD-LGVITD 85
DB 375 FVTNLTGKFNPAKFKIYEVDNQFVDSFTPTSKLQDVGTGQFDVIYSDNKTATVD 434

QY 86 LINPDIYRIDPNVLLDPADEKLEEBIQAPTSKSSKQQAQVVPVPMRKTEY-ISTEFNRY 144
DB 435 LLNGQS-----SSDKQYIIQVAYPDNS--STDNKGI-----DYLTEQNGKS 475

[illegible]

RESULT 9
D89824
hypothetical protein sdrD [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89824
F:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kohayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1385 <KUR>
A:Cross-references: UNIPROT:Q99W47; GB:BA000018; PID:gl3700454; PID:BAB41751.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: sdrD

	Query Match	8.3%	Score 230.5;	DB 2;	Length 1385;
	Best Local Similarity	20.4%	Pred. NO. 0.00023;		
	Matches 116;	Conservative	89;	Mismatches 236;	Indels 127; Gaps 20;
Qy	27	SGVVCRVKYCNLSL	IPDPKPFITYFPQNR	FVQYKATSLBKQHKHDL	LLTLPDGLGVTIDL 86
		::::	::::	::::	::::
Db	772	TGVI	-----NGA	NMTLDSGF--YKTPKYNLGNVWED	TNKGQKQDSTEKGISGVTVTL 823
Qy	87	INPD	-----	TYRID-----	PNVLDPDADEKLLBEEIQAP 115
		::::	::::	::::	::::
Db	824	KNENGEVLQTTKTI	KDGKYQPTGLENGYIKVFE	FETPSCGYTPTQVSGSTDEG-	IDNSGTST 882
Qy	116	TSSKRSQOHAKV	---VPMRKTEYISTEFNR	VGISNEKEPEVKIGVSKQOFTEEEIYK	170
		::::	::::	::::	::::
Db	893	TGVIKDKNDITD	SGFYKPTYNLGDYVWED	TNKGVDQDKBKGISGVTV	-----TLK 934
Qy	171	DRDSQITAEKTFHBAQKS	ISO---HYSKPRVTPEVMPFP	PFQMWINPFCQAVIFDSD	226
		::::	::::	::::	::::

Db 935 DENDKVLKTVTTDENGKQFTDLNNGTYKYVEFETPSGYTPT-----SVTSGN 981
 Qy 227 PAPKDTSGAALLEMMSQAMIRGMMDEG-----NQFVAYFLPVSETLKKRKRQDEE 277
 Db 982 DTEKDSNGLTTTGVVTKDA--DNMTLDSGFYKTPKYSILGDIYVWY-----DSNKGKQDSTE 1034
 Qy 278 EMDYAPDDVYDYKI-----AREYNMNVKNKASKGYEENYFFIFREBGGVYVYNEL 326
 Db 1035 K-----GIKQVKVILLNKEGVEIGTTKTDENGKYRFDNLDSGKYKVIKFTPTGL----- 1083
 Qy 327 ETRVRLSKRRKAGVQSGTGNALLVVKHRDMNNEKELEAQEARKAQLN-HEPEEEEEEME 385
 Db 1084 -----TOTGNTTDEDDKADGGEVDVTITDHDFTLDNGVYEETSDDSDSD 1129
 Qy 386 TEEKKAGSGDERQEKSGSSGEKESGSEHSGSESERE-EGDRDEASDKSGSGESESDEEA 444
 Db 1130 SD 1188
 Qy 445 RAARDKEEIFGSDANSEDDADSDDEDRGOAQGGSDNDSDSGSNGGQGRSRSHRSASPPP 504
 Db 1189 SD 1247
 Qy 505 SGSEHSAQEDGSEAAASDS-SEADSDSD 531
 Db 1248 SD 1275

 RESULT 10
 F90070
 Clumping factor B [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C:Accession: F90070
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogunuma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: F90070
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-877 <CUR>
 A:Cross-references: UNIPROT:Q99R07; GB:BA000018; PTD:gl3702588; PIDN:BAB43728.1; GSPDB:(S)
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: clfB

 Query Match 8.3%; Score 230; DB 2; Length 877;
 Best Local Similarity 22.6%; Pred. No. 0.00014;
 Matches 125; Conservative 76; Mismatches 207; Indels 145; Gaps 23;

 Qy 33 VKYCNSLPDIPF-DPKFITPPQNFVQYKATSLKQKHHDLLTEPDLGVVTDLIN--- 88
 Db 271 VDYSNSNMTPIADIK-----STNGDVAKAI-----YDILTCTFTFTDYNKKE 317
 Qy 89 -----PDTYRIDPNVLLDPADEKLEE-----ETDQPTSSKRS 121
 Db 318 NINGQFSLPFTDRAKAPKSGTYDANINI--ADEMFNKNITNYSSPIAGIDKPKNCANIS 375
 Qy 122 QOHAKVPMRKTEYISTEENRYGINSKEPEVKIGSVKQOFTTEIEIKYDRDSQITAE- 180
 Db 376 SQIIGVDTASGQNTYKQTVF-----VNPQKRVLGNTVWYIKGYQDKI--EESGSGKVSATDT 429
 Qy 181 --KTFE--DAQKSIQSHYSKPRVTPV-EVMPVPDFKMWINPCAQVIFDSDPAPKDTSCA 235
 Db 430 KLRIFEVNDTSKLSQSYADPNDSNLKEVTDQFNRIYIEHPNVASIKFGD----- 480
 Qy 236 AALEMMSQAMIRGMMDEEGNQFVAYFLPVSETLKKRKRQDEEEMDYAPDDVVDYKIAREY 295
 Db 481 --ITKTYVVLVEGHYDNTG-----KNLKTQVIQENVDPVTRDYSI---F 520
 Qy 236 NMNVKNKASKGYEENYFFIFREGDGVYVYNELTRVLRSKRRKAGVQSGTGNALLVVKHRD 355

434	Qy	--SCEDESSE-----	DEAARAARDEKEE	FGSDASDEDA	DSDD	EDRGOAQGSD	479
142	Db	EDNGDDSDDDNGD	ENGDAEDGDPAED	-----GDDAEDG	DGDDAEDGD	DAEDGDDA	199
480	Qy	NDSDSGNSGGGQR	SKSHSRAS	PPFGSGSEHAQEDG	SGFAAASD	SDSSEADSD	531
200	Db	EDGDAAEDGDAE	EDGDADGDGDAED	GDGDDNEDAE	DGDDAEDGDAE	DGDDAEDGDNE	251

Search completed: September 26, 2005, 13:55:35
Job time : 45 secs

GenCore version 5.1.1.6
(c) 1993 - 2005 CompuGen Ltd.

Copyright

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:35:20 ; Search time 167 Seconds
(without alignments)
1229.759 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPTTQTQQRDGRPNHS.....QEDGSEAAASDSEADSDSD 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseq_19808:*

2: Geneseq_19908:*

3: Geneseq_20008:*

4: Geneseq_20018:*

5: Geneseq_20028:*

6: Geneseq_20038:*

7: Geneseq_20038s:*

8: Geneseq_20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2764	100.0	531	2	AAY42226 Human pan
2	2764	100.0	531	7	ADD18712 Human dis
3	2764	100.0	531	8	ADO58688 Human reg
4	2764	100.0	531	8	ABM82102 Tumour-as
5	2744	99.3	531	4	AAB93517 Human pro
6	2658.5	96.2	552	4	ABG19682 Novel hum
7	2464	89.1	473	3	AAB56316 Human sec
8	1244.5	45.0	538	4	ABG19681 Novel hum
9	622	22.5	133	4	ABG19681 Novel hum
10	595	21.5	115	3	AAG03326 Human sec
11	283	10.2	478	4	ABG19412 Novel hum
12	283	10.2	478	8	AD512265 Human the
13	253	9.2	445	6	ABR53245 Protein s
14	253	9.2	445	7	ADK63670 Disease t
15	237.5	8.6	1633	6	ABU42513 Protein e
16	234.5	8.5	1802	3	AAY83170 Cell wall
17	234.5	8.5	1802	3	AAY70119 Staph. ep
18	234	8.5	552	6	ABU16533 Protein e
19	233.5	8.4	930	2	AAY08641 S. aureus
20	233.5	8.4	947	6	ABJ18940 Pathogen
21	233	8.4	932	4	AAY36845 Staphyloc
22	233	8.4	932	4	AAY34082 Staphyloc
23	232.5	8.4	995	6	ABM72437 Staphyloc
24	230.5	8.3	1385	6	ABU16400 Protein e
25	230	8.3	877	6	ABU42504 Protein e

ALIGNMENTS

RESULT 1

AAY42226

ID AAY42226 standard; protein; 531 AA.

XX AC AAY42226;

XX DT 20-DEC-1999 (first entry)

XX DE Human pancreatic differentiation 2 protein sequence.

XX KW Human; PD2; cancer; regulation; differentiation; neoplastic; therapy;

XX KW pancreatic differentiation 2; diagnosis; pancreatic adenocarcinoma;

XX KW phosphoprotein.

XX OS Homo sapiens.

XX FN WO9950408-A1.

XX PD 07-OCT-1999.

XX PF 26-MAR-1999; 99WO-US006633.

XX PR 27-MAR-1998; 98US-0079649P.

XX PA (UYNE-) UNIV NEBRASKA.

XX PI Batra SK, Hollingsworth MA;

XX DR WPI, 1999-591317/50.

XX N-PSDB; AAZ25433.

XX PT New phosphoprotein useful as targets for therapy of pancreatic

XX adenocarcinomas.

XX PS Claim 7; Fig 2; 97pp; English.

XX CC The present sequence is the human pancreatic differentiation 2 (PD2)

XX CC protein, which comprises an amino terminal helix-loop-helix domain and a

XX CC centrally localised nuclear transporter signal and nucleotide binding

XX CC site. The PD2 nucleotide sequence and a transformed host cell are useful

XX CC for screening a test compounds for PD2 modulating activity indicated by

XX CC an alteration in the phosphorylation of status of PD2. The host cells are

XX CC assessed for altered expression of pancreatic differentiation markers

XX CC (MUC-1 or carbonic anhydrase), and modulating activity is correlated with

XX CC an alteration in cellular morphology. The PD gene and protein represent

XX CC valuable targets in the differential diagnosis and therapy of pancreatic

XX CC adenocarcinomas

XX

Q	SQ	Sequence 531 AA;
		Query Match 100.0%; Score 2764; DB 2; Length 531; Best Local Similarity 100.0%; Pred. No. 1.3e-196; Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MAPTQTQORECHRPNSHRTLPERSGVGVCKVCNSLPDIPDPKFIYFPDQNRVQ 60
DB	1	MAPTQTQORECHRPNSHRTLPERSGVGVCKVCNSLPDIPDPKFIYFPDQNRVQ 60
QY	61	YKATSLKQKHDLTTEPDGLVTIDINPDTYRIDPNVLDPADKLEEEIQAPTSKR 120
DB	61	YKATSLKQKHDLTTEPDGLVTIDINPDTYRIDPNVLDPADKLEEEIQAPTSKR 120
QY	121	SQAHAKVPWMRKTEYISTEFNRYGISNEKEPEVKIGSVKQOFTSEIYKDSDSQAIAE 180
DB	121	SQAHAKVPWMRKTEYISTEFNRYGISNEKEPEVKIGSVKQOFTSEIYKDSDSQAIAE 180
QY	181	KTFEDAQKSISQHYSKPRVTPVEVMPFDFKWINPCAOVIFDSDPAPKTSGAAALEM 240
DB	181	KTFEDAQKSISQHYSKPRVTPVEVMPFDFKWINPCAOVIFDSDPAPKTSGAAALEM 240
QY	241	MSQAMIRGMDEEGNQVFAYFLPVEETLKKRKDOEEEMDYAPDDVDYDKIAREYNWVK 300
DB	241	MSQAMIRGMDEEGNQVFAYFLPVEETLKKRKDOEEEMDYAPDDVDYDKIAREYNWVK 300
QY	301	NKASKGYEENFYFFREGDGYYINELETRVRLSKRKAQVQSCTNALLVVKHRDMNEKE 360
DB	301	NKASKGYEENFYFFREGDGYYINELETRVRLSKRKAQVQSCTNALLVVKHRDMNEKE 360
QY	361	LEAQARKAQLLENHEPEEEEEETEKGAGSDERQEKGSSEKESGEHSSESER 420
DB	361	LEAQARKAQLLENHEPEEEEEETEKGAGSDERQEKGSSEKESGEHSSESER 420
QY	421	EGRDRDEASDKSGGDESEDSEDEARAARDKEEIFGSDADSDDEDKGOAQGGSDN 480
DB	421	EGRDRDEASDKSGGDESEDSEDEARAARDKEEIFGSDADSDDEDKGOAQGGSDN 480
QY	481	DSDSGSGGGQRSHRSRASPPFGSGSEHAQEDGSEAAAASDSEADSDSI 531
DB	481	DSDSGSGGGQRSHRSRASPPFGSGSEHAQEDGSEAAAASDSEADSDSI 531
RESULT 2		
ID	ADD18712	
XX	ADD18712 standard; protein; 531 AA.	
XX	AC	
XX	AC	
XX	15-JAN-2004 (first entry)	
XX	Human disease related protein SeqID143.	
DE	human; disease state; cytostatic; antiinflammatory; ophthalmological;	
XX	antiartherosclerotic; valvular; gene therapy;	
KW	hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;	
KW	inflammation; erythrocytosis; glycolysis; gluconeogenesis;	
KW	glucose transportation; catecholamine synthesis; iron transport;	
KW	nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;	
KW	retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;	
KW	inflammatory condition; wound healing.	
OS	Homo sapiens.	
XX	WO2003018621-A2.	
FN	06-MAR-2003.	
XX	23-AUG-2002; 2002WO-GB03892.	
XX	23-AUG-2001; 2001GB-00020558.	
FR	05-OCT-2001; 2001GB-00024037.	
XX		

AD058688
ID AD058688 standard; protein; 531 AA.
XX
AC ADO58688;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human regulatory molecule HRM-9.
XX
KW cytostatic; immunomodulator; agonist; antagonist; gene therapy;
KW human regulatory molecule; HRM; disease development; cell proliferation;
KW immune response; cancer.
XX
OS Homo sapiens.
XX
FN US2002058264-A1.
XX
PD 16-MAY-2002.
XX
PF 26-SEP-2001; 2001US-00840787.
XX
PR 23-SEP-1997; 97US-00933750.
PR 20-JAN-1999; 99US-00234613.
PR 03-MAR-2000; 2000US-00518665.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Hillman JL, Bandman O, Shah P, Au-Young J, Yue H;
PI Guegler KJ, Corley NC;
XX
XX WPI; 2004-459763/43.
DR N-PSDB; ADO58737.
XX
XX New human regulatory molecules, useful in the diagnosis and treatment of
XX cancer and immune disorders.
XX
XX Claim 1; SEQ ID NO 9; 115pp; English.
XX
XX The invention describes human regulatory molecules (HRM) (I) selected
XX from a group comprising the fully defined amino acid sequences of SEQ ID
XX NOS: 1-49. Also described are: an isolated polynucleotide (II) comprising
XX a nucleic acid sequence encoding (I) or the complement of the
XX polynucleotide (SEQ ID NOS:50-98); a composition comprising (II) and a
XX reporter molecule; an expression vector containing (II); a host cell
XX containing the vector; detecting (M1) expression of a nucleic acid in a
XX sample; screening (M2) a plurality of molecules to identify a ligand;
XX diagnosing (M3) a disease associated with gene expression in a sample
XX containing nucleic acids; a composition comprising (I) and a
XX pharmaceutical carrier for a labeling moiety; screening (M4) a plurality
XX of molecules to identify a ligand; preparation and purification of
XX antibodies; an antibody which specifically binds to (I); and detecting
XX protein expression in a sample. The new human regulatory protein
XX molecules which are expressed during disease development and the
XX polynucleotides which encode them satisfies a need in the art by
XX providing compositions which are useful in the diagnosis and treatment of
XX diseases associated with cell proliferation, particularly immune
XX responses and cancers. This is the amino acid sequence of a human
XX regulatory molecule.
XX
SQ Sequence 531 AA;
Query Match 100.0%; Score 2764; DB 8; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.3e-196;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPTIQQAQEDGHRPNSHRTLPSRSGVCRVKYCNLSLPIDPDPKFIITFPDQNRVQ 60
DB 1 MAPTIQQAQEDGHRPNSHRTLPSRSGVCRVKYCNLSLPIDPDPKFIITFPDQNRVQ 60
QY 61 YKATSLKQKHHDILTEPDGLGVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSKR 120
DB 61 YKATSLKQKHHDILTEPDGLGVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSKR 120

QY 121 SQQAKVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEIYKDRDSQITAE 180
DB 121 SQQAKVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIHQHYSKPRVTVPVEVMPVFPDFKQWNPFCQVIFDSDPAPKD7SGAAALEM 240
DB 181 KTFEDAQKSIHQHYSKPRVTVPVEVMPVFPDFKQWNPFCQVIFDSDPAPKD7SGAAALEM 240
QY 241 MSQAMIRGMMDERGNQFVAYFLPVEETLKQRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
DB 241 MSQAMIRGMMDERGNQFVAYFLPVEETLKQRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
QY 301 NKASKGYEENYFFIPREGDGVYVYNELETRVRLSKRAKAGVQSGTALLVVHHRDMNEKE 360
DB 301 NKASKGYEENYFFIPREGDGVYVYNELETRVRLSKRAKAGVQSGTALLVVHHRDMNEKE 360
QY 361 LEAQEARKAQLENHPEPEEEEEEMETEKEAGGSDEEKEKGSSEKESGSDHSGSESER 420
DB 361 LEAQEARKAQLENHPEPEEEEEEMETEKEAGGSDEEKEKGSSEKESGSDHSGSESER 420
QY 421 EECGRDEASDKSGSDESEDEARAAARDKEEIEFGSDADSEDDADSDDEDRGQAQGGSDN 480
DB 421 EECGRDEASDKSGSDESEDEARAAARDKEEIEFGSDADSEDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGNGGGQSRSHRSRSGASPPSGSEHSAQSDGSEAAASDSEADSDSD 531
DB 481 DSDSGNGGGQSRSHRSRSGASPPSGSEHSAQSDGSEAAASDSEADSDSD 531
RESULT 4
ABM82102
ID ABM82102 standard; protein; 531 AA.
XX
AC ABM82102;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO83014, SEQ:5424.
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.
XX
XX 02-OCT-2002; 2002US-0414971P.
XX
XX (GETH) GENENTECH INC.
XX
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
DR N-PSDB; ACN40565.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 5424; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are

overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT polypeptide of the invention

Sequence 531 AA;

Query Match 100.0%; Score 2764; DB 8; Length 531;
 Best Local Similarity 100.0%; Pred. No. 1.3e-196;
 Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60
 Db 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Qy 61 YKATSLKQKHDLTLPDGLVTTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
 Db 61 YKATSLKQKHDLTLPDGLVTTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 120

Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGVSVKQQTTEEIYKDRDSQITAE 180
 Db 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGVSVKQQTTEEIYKDRDSQITAE 180

Qy 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKDTSGAAALEM 240
 Db 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKDTSGAAALEM 240

Qy 241 MSQAMTIGWMDSEGNQVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
 Db 241 MSQAMTIGWMDSEGNQVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300

Qy 301 NKASKGYEENYFTIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDNEKE 360
 Db 301 NKASKGYEENYFTIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDNEKE 360

Qy 361 LEAQEARKAOLENHEPEEEEEEMETEKEAGGSDEEKEKGSSEKEGSEDEHSGSESR 420
 Db 361 LEAQEARKAOLENHEPEEEEEEMETEKEAGGSDEEKEKGSSEKEGSEDEHSGSESR 420

Qy 421 EEGDRDASPKSGGSEDESSEDEARAKDEEITFGSDADSEDDADSDDEDRGAQGGSDN 480
 Db 421 EEGDRDASPKSGGSEDESSEDEARAKDEEITFGSDADSEDDADSDDEDRGAQGGSDN 480

Qy 481 DSDSGSGGGQGRSHRSRSPFPSSGSEHAQEDGSEAAASDSEADSDSD 531
 Db 481 DSDSGSGGGQGRSHRSRSPFPSSGSEHAQEDGSEAAASDSEADSDSD 531

RESULT 5

ID AAB93517

XX AAB93517 standard; protein; 531 AA.

AC AAB93517;

XX 26-JUN-2001 (first entry)

DT Human protein sequence SEQ ID NO:12853.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

XX

XX

XX

XX

XX

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 12853; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dr primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

XX present invention

XX Sequence 531 AA;

XX Query Match 99.3%; Score 2744; DB 4; Length 531;

XX Best Local Similarity 99.4%; Pred. No. 3.9e-195;

XX Matches 528; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Db 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Qy 61 YKATSLKQKHDLTLPDGLVTTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 120

Db 61 YKATSLKQKHDLTLPDGLVTTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 120

Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGVSVKQQTTEEIYKDRDSQITAE 180

Db 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGVSVKQQTTEEIYKDRDSQITAE 180

Qy 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKDTSGAAALEM 240

Db 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKDTSGAAALEM 240

QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDDEEMDYAPDDVYDKIAREYNNVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDDEEMDYAPDDVYDKIAREYNNVK 300
QY 301 NKASKGYENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAQLNHEPEEHEEEMETEKEAGGSDDEOEKSGSSEKEGSEDEHSGSSESER 420
DB 361 LEAQEARKAQLNHEPEEHEEEMETEKEAGGSDDEOEKSGSSEKEGSEDEHSGSSESER 420
QY 421 EGRDRDEASDKSGGSEDESSEDEARAARDKEEIFGSDADSDDDDDRQAGGSDN 480
DB 421 EGRDRDEASDKSGGSEDESSEDEARAARDKEEIFGSDADSDDDDDRQAGGSDN 480
QY 481 DSDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 531
DB 481 DSDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 531

RESULT 6
ABG19682
ID ABG19682 standard; protein; 553 AA.
XX AC ABG19682;
XX XX
DT 18-FEB-2002 (first entry)
XX XX
DE Novel human diagnostic protein #19673.
XX XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX XX
PN WO200175067-A2.
XX XX
PD 11-OCT-2001.
XX XX
PF 30-MAR-2001; 2001WO-US0008631.
XX XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX XX
FA (HYSE-) HYSEQ INC.
XX XX
PI Drmanac RT, Liu C, Tang YT;
XX XX
DR WPI; 2001-639362/73.
DR N-FSDB; AAS83869.
XX XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX XX
PS Claim 20; SEQ ID NO 50041; 103pp; English.
XX XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 553 AA;

Query Match 96.2%; Score 2658.5; DB 4; Length 553;
Best Local Similarity 95.04; Pred. No. 9.2e-189;
Matches 515; Conservative 6; Mismatches 10; Indels 11; Gaps 1;

QY 1 MAPIQTQAOQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNFVQ 60
DB 12 MAPIQTQAOQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNFVQ 71
QY 61 YKATSLKQKHDLLEPDLGVITDILNPDTYRIDPNVLLDPADEKLEBEIQAPTSSKR 120
DB 72 YKATSLKQKHDLLEPDLGVITDILNPDTYRIDPNVLLDPADEKLEBEIQAPTSSKR 131
QY 121 SQQHAVVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEEIIYKDRDSQITAE 180
DB 132 SQQHAVVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEEIIYKDRDSQITAE 191
QY 181 KTFEDAQKS-----ISQHSKPRVTPVEVMVFPDFKWINPQCAQVIFDSPPAP 229
DB 192 KTFEDAQKSVEGLGWGEARISQHSKPRVTPVEVMVFPDFKWINPQCAQVIFDSPPAP 251
QY 230 KDTSGAAALEMMSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQSEEMDYAPDDVYDY 289
DB 252 KDTSGAAALEMMSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQSEEMDYAPDDVYDY 311
QY 290 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL 349
DB 312 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL 371
QY 350 VVXHRDMEKELEAQEARKAQLNHEPEEHEEEMETEKEAGGSDDEOEKSGSSEKEGS 409
DB 372 VVXHRDMEKELEAQEARKAQLNHEPEEHEEEMETEKEAGGSDDEOEKSGSSEKEGS 431
QY 410 EDEHSGSESEREGDRDEASDKSGGSEDESSEDEARAARDKEEIFGSDADSDDDADSDDE 469
DB 432 EDEHSGSESEREGDRDEASDKSGGSDSDSYKARAARDKEEIFGSDADSDDDADSDDE 491
QY 470 DRGQAQGGSDNDSGNSGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 529
DB 492 DRGQAQGGSDNDSGNSGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 551
QY 530 SD 531
DB 552 SD 553

RESULT 7
AAB56316
ID AAB56316 standard; protein; 473 AA.
XX AC AAB56316;
XX XX
DT 13-MAR-2001 (first entry)
XX XX
DE Human secreted protein sequence encoded by gene 106 SEQ ID NO:410.
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;

KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative.
 XX Homo sapiens.
 XX WO200070042-A1.
 XX 23-NOV-2000.
 XX 11-MAY-2000; 2000WO-US012788.
 XX 13-MAY-1999; 99US-0134068P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
 PI Duan RD, Florence KA, Soppet DR;
 XX WPI; 2000-679828/66.
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX Disclosure; Page 1041-1042; 1065pp; English.
 XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention
 XX SQ Sequence 473 AA;
 Query Match 89.1%; Score 2464; DB 3; Length 473;
 Best Local Similarity 99.8%; Pred. No. 2.1e-174;
 Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MAPTIQQAQREDGHRNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYPPDQNRVQ 60
 Db 1 MAPTIQQAQREDGHRNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYPPDQNRVQ 60
 Qy 61 YKATSLKQKHGDLTTPDILGVITDLINPDTYRIDPNVLDPADEKLEEEICAPTSSKR 120
 Db 61 YKATSLKQKHGDLTTPDILGVITDLINPDTYRIDPNVLDPADEKLEEEICAPTSSKR 120
 Qy 121 SQQAKVVPWNRKTEYISTFNRYGISNEKPEVKIGVSKQPTTEBEIYKDRDSQITAE 180
 Db 121 SQQAKVVPWNRKTEYISTFNRYGISNEKPEVKIGVSKQPTTEBEIYKDRDSQITAE 180
 Qy 181 KTFEDAQKSIQSHKSKPRVTFVEMVFPDFKWINPCAQVIFDSDPAPKDTSGAAALEM 240
 Db 181 KTFEDAQKSIQSHKSKPRVTFVEMVFPDFKWINPCAQVIFDSDPAPKDTSGAAALEM 240
 Qy 241 MSQAMIRGMDEENQVAFVLPVEETLKKRRDQEEEMDYAPDDVYDKIAREYNWVK 300

Db 241 MSQAMIRGMDEENQVAFVLPVEETLKKRRDQEEEMDYAPDDVYDKIAREYNWVK 300
 Qy 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVSGTNNALLVWHRDMNEKE 360
 Db 301 NKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRAKAGVSGTNNALLVWHRDMNEKE 360
 Qy 361 LEAQARQAQLENHEPEEEEEEMETEKEAGSGDEEOKGSSSEKEGSEDEHSGSESER 420
 Db 361 LEAQARQAQLENHEPEEEEEEMETEKEAGSGDEEOKGSSSEKEGSEDEHSGSESER 420
 Qy 421 EREGDRDEASDKSGSGEDESSEDEARAARDKEIFGSDADSEDDADSDDDDRGQ 473
 Db 421 EREGDRDEASDKSGSGEDESSEDEARAARDKEIFGSDADSEDDADSDDDDRGQ 473
 RESULT 8
 ABB59163
 ID ABB59163 standard; protein; 538 AA.
 XX ABB59163;
 AC ABB59163;
 XX 26-MAR-2002 (first entry)
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 4281.
 DE Drosophila melanogaster polypeptide SEQ ID NO 4281.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEXE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 WPI; 2001-656860/75.
 DR N-PSDB; ABL03266.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 4281; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 538 AA;
 Query Match 45.0%; Score 1244.5; DB 4; Length 538;
 Best Local Similarity 50.0%; Pred. No. 9.3e-84;
 Matches 271; Conservative 66; Mismatches 172; Indels 39; Gaps 11;
 Qy 1 MAPTIQQAQREDGHRNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIYPPDQNRVQ 60
 Db 1 MPPTINNSAVNSAEK-RPQRTKRSKEICRVKGNLDPDIPDLKFLQKFPDHSRFRVQ 59

27-AUG-2003.
20-DEC-2002; 2002EP-00102902.
20-DEC-2001; 2001EP-00130253.
(CELL-) CELLZONE AG.
Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
Michon A, Leutwein C, Rick J;
WPI; 2003-638460/61.
N-PSDB; ADK63671.
New proteins and protein complexes from eukaryotes, useful as targets in drug screening, or in diagnosing or screening for the presence of a disease or disorder, or a predisposition for developing a disease or disorder in a subject.
Disclosure; SEQ ID NO 1635; 13pp; English.
The invention relates to novel protein complexes comprising a first and a second protein, or its derivative, fragment, homologue or variant. The proteins are selected from given protein complexes, which are not defined in the specification. The variants are encoded by nucleic acids that hybridize to the nucleic acids encoding the proteins under low stringency conditions. The protein complexes are useful as targets for an active agent of a pharmaceutical. These protein complexes are particularly useful as drug targets for the treatment or preventing of a disease or disorder. The complexes and methods above are useful in diagnosing or screening for the presence of a disease or disorder or a predisposition for developing a disease or disorder in a subject. These are also useful in screening for a drug for treatment or prevention of a disease or disorder. The molecule that modulates the amount, activity or protein components of the complex is useful for the manufacture of a medicament for the treatment or prevention of a disease or disorder. This sequence corresponds to a protein of the invention. (Note: the sequence data for this patent did not form part of the printed specification but was obtained from the EPO in electronic format).
Sequence 445 AA;
Query Match 9.28; Score 253; DB 7; Length 445;
Best Local Similarity 22.18; Pred. No. 3.2e-10;
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;
23 LPERSGVVCVKYCNLSLPDIPDPKFTYPT-----FQNRVQVKATSLKQHK 71
1 MSKQEVIAPIKQNSLPVQLPKLVYPSPEPTNADSSQLNSLYIKTNVNLIQ-- 58
72 HDLLEPDLGVTIDLI-----NPDYRIDPNVLLDPADEKLEELIEIQAPTSKRS 121
59 -----DEDLGMPVLMKFPGLLNKLDKLLYGFD-NVKLDKDRILLRD----PRIDRLT 108
122 QOHAKVVPWRKTEVISTEFNRYGISNEKEPVKIGSVKQFTEEEYKDRDSQ----- 175
109 KTDISKVTFLETRTEVTSNTIAHDNTSLKRRRL-----DGDSDDENLDV 154
176 ---ITALEKTFEQAQKISQHSYSPRTVPVEVMPVFPDFKWINPCACQVIFDSDPAPKDT 232
155 NHIISRVGEFTNKTDK--WQHPVKGVKVMVKWDLDP-----TASMDQVVF-----ILKF 203
233 SGAAALEMMSOAMIR-GM---MDEGNOFVAYFLPVEETLKRRQDOEEMDVAPDDVD 288
204 MGSASLTKEKSLNTGIFRPVELEDEWISWATDKDSAILLENELEKGMDEMDDSHE 263
289 ---YKIAREYNWVKKASKGYEENYFFIFREGDGV--YNELETRVLSRRKAKGVQSG 344
264 GKIKYKRIIRDYDKQVAKPMT-LAIRLNDKDGIAIYKPKSLKELRRRRVNDLIK- 321
345 TNALLVVKH-----RDMNEKELEAQBARKAQLEN-----HEPEEEBEMEETEEK 389
322 ----LVKEHDIDOLNVTLRNPSTKEANIRDKLRMKFDPINFATVDEDEDEEQPEDVKK 377
390 EAGGSDEFEQKSGSSKESGSEDEHSGSESEREGGDRDEADSKSGSGEDESSEDEARAARD 449
378 ESEG--DSKTEGSEQEGENKDEEIKQEKENEQ-----DEENKQDENRAADT 422
450 KEEIFGSDADSDSDADSDDDDRGQAQ 475
423 PET---SDAVHTQKPEEKETLOEE 445
RESULT 15
ABU42513
ID ABU42513 standard; protein; 1633 AA.
XX
AC ABU42513;
XX
XX 19-JUN-2003 (first entry)
DT
XX Protein encoded by Prokaryotic essential gene #28040.
DE
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Staphylococcus epidermidis.
OS
XX WO200277183-A2.
PN
XX 03-OCT-2002.
PD
XX 21-MAR-2002; 2002WO-US009107.
PF
XX 21-MAR-2001; 2001US-00815242.
PR
XX 06-SEP-2001; 2001US-00948993.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX WPI; 2003-029926/02.
DR N-PSDB; ACA46383.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
PT
XX Claim 25; SEQ ID NO 70437; 1766pp; English.
PS
XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of the strains; or (13) identifying the target of a compound that inhibits proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ

SQ		Sequence 1633 AA;
Query Match		8.6%; Score 237.5; DB 6; Length 1633;
Best Local Similarity		21.2%; Pred. No. 2.5e-08;
Matches 131; Conservative		85; Mismatches 238; Indels 165; Gaps 25;
QY	48	FTYPPDONRFVQYKATSLKOHKDLLETPDLGVTIDLINPDYRIDP-----NVLLDP 102
DB	707	YVTLKDSNNRELQRYVTTDOSGHYQFDNLQNGT--YTFEPAIPDNTTPSPANNSTNDADS 764
QY	103	ADEKLEEEIQAPTSTSRQOHAKV-----VPMRKTEYISTEFNRYGI--SNEKPEVK 154
DB	765	DGERDGRKVVVAKGTINNADNMVDTGFLTPKYNVGDYVWEDTNKOGIQDDNEKGISN 824
QY	155	IGVSVVKQ-----QFTEEEIY-----KDRDSQ 175
DB	825	VKVTILKNKNGDTIGTITTSNGKYEFTGLENGDYTFETPEGYTPTKQNSGSDGKDSN 884
QY	176	ITAIKTFEDA-QKISQHSYKPRVTPVPMVFPDFKMWNPQCAQVIFDSDP----- 227
DB	885	GTKTTTVTKADNRKIDSGFYKPIYN-----LGDI-VWEDTNKOGIQDSEKGISGVK 936
QY	228	-APKDTSGAA-----ALEMWSQAMIGMDEGNQFVAVFLPVEETLKKRKRQEEEMDY- 281
DB	937	VTLKDNKNGNAIGTITTDASGHYQPKGL--ENGYTFEFTPSGYTPTKANSQDITVDSN 994
QY	282	-----APDRYVD---YKIAR-----EYNNVNKVK-----ASKGY----- 307
DB	995	GITTTGIINGADNLITDSGFYKTPKYVGDYVWEDTNKOGIQDDNEKGISGVKVTLKDEK 1054
QY	308	-----EENYHIFREGD-GVYVNELETRVLSKRAKAGVQSGTNALLVVKHRDMN 357
DB	1055	GNIISTTTTDENGRYQFDNLDSGNVYIIHFKEKPEGMTQTITANSG-----NDD 1100
QY	358	EKELEAQEARKAQLENNH-----EPEEEEEEMETEKEAGGSD 395
DB	1101	EKDADGEDVR-VTIIDHDDFSDNGYFDDSDSDSDADSDSDSDSDSDSDSDSDSDSD 1159
QY	396	EEQKSGSSEKESDEHSGSESERE-EGDRDEASDK-SGSGEDESSEDEAARAADKEEI 453
DB	1160	SD 1219
QY	454	FGSDADSEDDADSDIEDRQAGGSDNDSDSGSGGQSRSHRSASFPFSGSEHSAQE 513
DB	1220	SDSDADSDSDADSDSDADSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 1276
QY	514	DGSEAAADS-SEADSDD 531
DB	1277	DADSDDSDSDSDADSDSD 1295

Search completed: September 26, 2005, 13:51:43
Job time : 171 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:36:10 ; Search time 179 Seconds
(without alignments)
1519.073 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPTIQTQAREDRHRPNH.....QEDGSEAAASDSSSEADSDSD 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2764	100.0	531	2 Q9H166	Q9H166 homo sapien
2	2744	99.3	531	2 Q9NUU9	Q9NUU9 homo sapien
3	2718	98.3	535	2 Q8K2T8	Q8K2T8 mus musculus
4	2708	98.0	535	2 Q9J199	Q9J199 mus musculus
5	2531.5	91.6	510	2 Q75239	Q75239 homo sapien
6	2183.5	79.0	520	2 Q6F2Y1	Q6F2Y1 xenopus tro
7	1995	72.2	485	2 Q8N7H5	Q8N7H5 homo sapien
8	1984	71.8	377	2 Q9CS63	Q9CS63 mus musculus
9	1935	70.0	407	2 Q68F51	Q68F51 xenopus lae
10	1244.5	45.0	538	2 Q9VN55	Q9VN55 drosophila
11	1129	40.8	468	2 Q7PXA3	Q7PXA3 anopheles g
12	645	23.3	425	2 P90783	P90783 caenorhabdi
13	372.5	13.5	589	2 Q8RW91	Q8RW91 arabidopsis
14	361.5	13.1	593	2 Q9MA04	Q9MA04 arabidopsis
15	358.5	13.0	451	2 Q6ZD92	Q6ZD92 oryza sativ
16	335.5	12.1	547	2 Q9CA82	Q9CA82 arabidopsis
17	326.5	10.4	386	2 Q6C509	Q6C509 yarrowia li
18	280.5	10.1	457	2 Q9US06	Q9US06 schizosacch
19	265.5	9.6	791	2 Q2DGL1	Q2DGL1 figu rubrip
20	257.5	9.3	571	2 Q8MTN7	Q8MTN7 trichinella
21	253	9.2	445	1 PAF1_YEAST	P38351 saccharomyc
22	251.5	9.1	538	2 Q9ET15	Q9ET15 mus musculus
23	250.5	9.1	458	2 Q6BT93	Q6BT93 dbaryomyce
24	245.5	8.9	784	2 Q7LZ90	Q7LZ90 torpedo cal
25	245	8.9	1848	2 Q7RGP8	Q7RGP8 bos taurus
26	244.5	8.8	934	2 Q9GMD3	Q9GMD3 bos taurus
27	243.5	8.8	438	2 Q6FXJ9	Q6FXJ9 candida gla
28	241	8.7	792	2 Q9YTL7	Q9YTL7 ateline her
29	240	8.7	1394	1 CNG4_BOVIN	Q2B181 bos taurus
30	238.5	8.6	946	2 Q7QC53	Q7QC53 anopheles g
31	238	8.6	613	2 Q6UDM5	Q6UDM5 brachydanio

32	236.5	8.6	1633	2	Q8CMP4	Q8CMP4 staphylococ
33	235	8.5	1451	2	Q8I2D8	Q8I2D8 plasmodium
34	234.5	8.5	1733	2	Q9KI14	Q9KI14 staphylococ
35	234	8.5	498	2	Q8MTN8	Q8MTN8 trichinella
36	234	8.5	953	2	Q9NW48	Q9NW48 staphylococ
37	234	8.5	953	2	Q7A781	Q7A781 staphylococ
38	234	8.5	1323	2	Q9NB35	Q9NB35 plasmodium
39	233.5	8.4	947	2	Q86487	Q86487 staphylococ
40	233	8.4	955	2	Q8NXX7	Q8NXX7 staphylococ
41	232	8.4	957	2	Q6GBS6	Q6GBS6 staphylococ
42	231.5	8.4	873	2	Q6GDH2	Q6GDH2 staphylococ
43	230.5	8.3	843	2	Q6GLM0	Q6GLM0 xenopus lae
44	230.5	8.3	897	2	Q13098	Q13098 xenopus lae
45	230.5	8.3	1385	2	Q99W47	Q99W47 staphylococ

ALIGNMENTS

RESULT 1

Q9H166 ID Q9H166 PRELIMINARY; PRT; 531 AA.
AC Q9H166;
DT- 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein PD2.
GN Name=PD2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Batra S.K., Choudhury A., Keita M., Schmed B.M., Vanlith M.
RA Walter N.A.R., Jorkest J., Sikela J.M., Metzgar R.S.,
RA Hollingsworth M.A.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.S.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ401156; CAC20564.1; -.

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE RIKEN cDNA 5730511K23.
 GN Name=5730511K23Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg E., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Schetz T.F.,
 RA Brownstein M.J., Udwin R.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC029843; AAH29843.1; -;
 DR MGD; MGI:1923988; 5730511K23Rik.
 DR InterPro; IPR007133; Pfam.
 DR Pfam; PF03985; Pfam; 1.
 SQ SEQUENCE 535 AA; 60534 MW; 7A5EAB1284988070 CRC64;

 Query Match 98.13%; Score 2718; DB 2; Length 535;
 Best Local Similarity 98.11%; Pred. No. 7,1e-121;
 Matches 525; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

 QY 1 MAPTTQQAQREDGHRPNSHRTLPSRGVVCVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
 DB 1 MAPTTQQAQREDGHRPNSHRTLPSRGVVCVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
 QY 61 YKATSLKQKHDLITLTPDLGVITDILNPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
 DB 61 YKATSLKQKHDLITLTPDLGVITDILNPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
 QY 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
 DB 61 YKATSLKQKHDLITLTPDLGVITDILNPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
 QY 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
 QY 181 KTFEDAQKSIQSHYKSKPRVTPVEVMPVPDFQWNPVPCAVIFDSDPAPKDTSGAAALEM 240
 DB 181 KTFEDAQKSIQSHYKSKPRVTPVEVMPVPDFQWNPVPCAVIFDSDPAPKDTSGAAALEM 240
 QY 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVDYDKIAREYNVNVK 300
 DB 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVDYDKIAREYNVNVK 300
 QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
 DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
 QY 361 LEAQEARKAQLENHPEEEEEETETKEAGGSDDEQEGSSSEKSGSEDEHSGSESR 420
 DB 361 LEAQEARKAQLENHPEEEEEETETKEAGGSDDEQEGSSSEKSGSEDEHSGSESR 420
 QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480
 DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480

DB 361 LEAQEARKAQLENHPEEEEEETETKEAGGSDDEQEGSSSEKSGSEDEHSGSESR 420
 QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480
 DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480
 QY 481 DSDSGSNGGQGR-----SRSHRSASPPPSGSEHSAQEDGSEAAAAASDSSEADSDSD 531
 DB 481 DSDSGSDGGQGRSRQSRSRSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 535

 RESULT 4
 Q9J999
 ID Q9J999 PRELIMINARY; PRT; 535 AA.
 AC Q9J999;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Mus musculus brain cDNA, clone MNCb-6444, similar to Homo sapiens cDNA
 FLJ11123, clone PLACE1006167.
 GN Name=5730511K23Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL;
 RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
 RA Hashimoto K.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB041615; BAA95098.1; -;
 DR MGD; MGI:1923988; 5730511K23Rik.
 DR InterPro; IPR007133; Pfam.
 DR Pfam; PF03985; Pfam; 1.
 SQ SEQUENCE 535 AA; 60534 MW; 6D7EBB1ECDC8C075 CRC64;

 Query Match 98.0%; Score 2708; DB 2; Length 535;
 Best Local Similarity 97.9%; Pred. No. 2,1e-120;
 Matches 524; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

 QY 1 MAPTTQQAQREDGHRPNSHRTLPSRGVVCVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
 DB 1 MAPTTQQAQREDGHRPNSHRTLPSRGVVCVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
 QY 61 YKATSLKQKHDLITLTPDLGVITDILNPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
 DB 61 YKATSLKQKHDLITLTPDLGVITDILNPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
 QY 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
 DB 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTETEEIYKDRDSQITAE 180
 QY 181 KTFEDAQKSIQSHYKSKPRVTPVEVMPVPDFQWNPVPCAVIFDSDPAPKDTSGAAALEM 240
 DB 181 KTFEDAQKSIQSHYKSKPRVTPVEVMPVPDFQWNPVPCAVIFDSDPAPKDTSGAAALEM 240
 QY 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVDYDKIAREYNVNVK 300
 DB 241 MSQAMIRGMWDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVDYDKIAREYNVNVK 300
 QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
 DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
 QY 361 LEAQEARKAQLENHPEEEEEETETKEAGGSDDEQEGSSSEKSGSEDEHSGSESR 420
 DB 361 LEAQEARKAQLENHPEEEEEETETKEAGGSDDEQEGSSSEKSGSEDEHSGSESR 420
 QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480
 DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDADSDDEDEFGAHRGSDN 480

```
Qy 481 DSDSGSGGCGOR-----SRSHRSRSPFSGSEHSAQEDGSEAAASDSSEADSDSD 531
Db 481 DSDSGSDGGGCGORSGSRSRSPFSGSEHSAQEDGSEAAASDSSEADSDSD 535

RESULT 5
O75239 PRELIMINARY; PRT; 510 AA.
AC O75239
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F23149.1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., De L., Regala W., Terry A., Gaines J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Irons S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005239; AAC2503.1; -.
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam; 1.
SQ SEQUENCE 510 AA; 57456 MW; CACE73EDC7290CE8 CRC64;

Query Match 91.6%; Score 2531.5; DB 2; Length 510;
Best Local Similarity 91.7%; Pred. No. 4.4e-112;
Matches 498; Conservative 0; Mismatches 0; Indels 45; Gaps 3;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTTYPDQNRVQ 60
Db 1 MAPIQTQAQREDG-----RSGVVCRVKYCNLSLPDIPDPKFTTYPDQNRVQ 50

Qy 61 YKATSLKQKHDLTPEDLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 51 YKATSLKQKHDLTPEDLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSKR 110

Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180
Db 111 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 170

Qy 181 KTFEDAQKSIQHSYKPRVTPEVMPVFPDFKWINPCAQVIFSDPAPKDTSGAAALEM 240
Db 171 KTFEDAQKSIQHSYKPRVTPEVMPVFPDFKWINPCAQVIFSDPAPKDTSGAAALEM 207

Qy 241 MSQAMIRGMDDEEGNOFVAYFLPVEETLKRRKRDQEEEMDYPDDVDYVYKILAREYNWVK 300
Db 208 MSQAMIRGMDDEEGNOFVAYFLPVEETLKRRKRDQEEEMDYPDDVDYVYKILAREYNWVK 267

Qy 301 NKASKGYEENYFTIFREGDGVYVNELETR-----VRLSKRAKAGVQSGTNAL 348
Db 268 NKASKGYEENYFTIFREGDGVYVNELETRYSAHSYLSLDLVRLSKRAKAGVQSGTNAL 327

Qy 349 LVYKHRDMKEKELEQAKQAQLENHEPEEEEEETETEKEAGGSDERQEKSGSSEKEG 408
Db 328 LVYKHRDMKEKELEQAKQAQLENHEPEEEEEETETEKEAGGSDERQEKSGSSEKEG 387

Qy 409 SEDEHSGSSESEEGDRDASDCKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSD 468
Db 388 SEDEHSGSSESEEGDRDASDCKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSD 447

Qy 469 EDGQAQGGSDNDSGSGNGGQGRSHRSRSPFSGSEHSAQEDGSEAAASDSSEADS 528
Db 448 EDGQAQGGSDNDSGSGNGGQGRSHRSRSPFSGSEHSAQEDGSEAAASDSSEADS 507
```

```
Qy 529 DSD 531
Db 508 DSD 510

RESULT 6
Q6P2Y1 PRELIMINARY; PRT; 520 AA.
AC Q6P2Y1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein MGC76249.
GN Name=MGC76249;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Schmutz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC064253; AAH64253.1; -.
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam; 1.
KW Hypothetical protein.
SQ SEQUENCE 520 AA; 59064 MW; 76D526C423C459A7 CRC64;

Query Match 79.0%; Score 2183.5; DB 2; Length 520;
Best Local Similarity 80.2%; Pred. No. 1.2e-95;
Matches 429; Conservative 40; Mismatches 45; Indels 21; Gaps 8;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTTYPDQNRVQ 60
Db 1 MAPIQTQAQREDGHRSSSHRTLPERSGVVCRVKYCNLTLPDIPDPKFTTYPDQNRVQ 60

Qy 61 YKATSLKQKHDLTPEDLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 61 YKATSLKQKHDLTPEDLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSKR 120

Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180
Db 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180

Qy 181 KTFEDAQKSIQHSYKPRVTPEVMPVFPDFKWINPCAQVIFSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQHSYKPRVTPEVMPVFPDFKWINPCAQVIFSDPAPKDTSGAAALEM 240
```



```
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kohno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tegami M., Teraoka Y., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL EMBL; AK017762; BAB30513.1; -;
DR MGD; MGI:1923988; 5730511K2Rik.
DR InterPro; IPR0071133; Pfam.
DR Pfam; PF03985; Pfam; 1.
DR Hypothetical protein.
KT NON TER 377
SQ SEQUENCE 377 AA; 43836 MW; 4CE500D24EF5CEA CRC64;

Query Match 71.8%; Score 1984; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 2.3e-86;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPDQNRVQ 60
Db |||||
1 MAPIQTQQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPDQNRVQ 60

QY 61 YKATSLKQKHGDLITPEPLGVITIDLPNTYRIDPNTLVLLDPADEKLEEEIQAPTSSKR 120
Db |||||
61 YKATSLKQKHGDLITPEPLGVITIDLPNTYRIDPNTLVLLDPADEKLEEEIQAPTSSKR 120

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGVSKQOFTBEIYKDRDSQITATE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGVSKQOFTBEIYKDRDSQITATE 180

QY 181 KTFEDAQKTSIQHYSKPRVTPEVMPVFPDFKQWNPVCAQVIFDSDPAKPTSGAAALEM 240
Db |||||
181 KTFEDAQKTSIQHYSKPRVTPEVMPVFPDFKQWNPVCAQVIFDSDPAKPTSGAAALEM 240

QY 241 MSQAMIRGMDDEEGNQFVAYFLPVEETLKKRKRDQEEENDYAPDDVYDYKIAREYNWVK 300
Db |||||
241 MSQAMIRGMDDEEGNQFVAYFLPVEETLKKRKRDQEEENDYAPDDVYDYKIAREYNWVK 300

QY 301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360
Db |||||
301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360

QY 361 LEAQEARKAQLNHEPE 377
Db |||||
361 LEAQEARKAQLNHEPE 377

RESULT 9
Q68F51 PRELIMINARY; PRT; 407 AA.
ID Q68F51
AC Q68F51;
DT 25-OCT-2004 (T-EMBLrel. 28, Created)
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DE LOC446278 protein (Fragment).
GN Names:LOC446278;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
```

```
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
EX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshitoki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RL Klein S., Gerhard D.S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC079993; AAH79993.1; -;
DR InterPro; IPR0071133; Pfam.
DR Pfam; PF03985; Pfam; 1.
FT NON TER 407
SQ SEQUENCE 407 AA; 47154 MW; 6CE32A7307186F83 CRC64;

Query Match 70.0%; Score 1935; DB 2; Length 407;
Best Local Similarity 90.0%; Pred. No. 5.2e-84;
Matches 367; Conservative 22; Mismatches 15; Indels 4; Gaps 2;

QY 1 MAPIQTQQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPDQNRVQ 60
Db |||||
1 MAPIQTQQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPDQNRVQ 60

QY 61 YKATSLKQKHGDLITPEPLGVITIDLPNTYRIDPNTLVLLDPADEKLEEEIQAPTSSKR 120
Db |||||
61 YKATSLKQKHGDLITPEPLGVITIDLPNTYRIDPNTLVLLDPADEKLEEEIQAPTSSKR 120

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGVSKQOFTBEIYKDRDSQITATE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGVSKQOFTBEIYKDRDSQITATE 180

QY 181 KTFEDAQKTSIQHYSKPRVTPEVMPVFPDFKQWNPVCAQVIFDSDPAKPTSGAAALEM 240
Db |||||
181 KTFEDAQKTSIQHYSKPRVTPEVMPVFPDFKQWNPVCAQVIFDSDPAKPTSGAAALEM 240

QY 241 MSQAMIRGMDDEEGNQFVAYFLPVEETLKKRKRDQEEENDYAPDDVYDYKIAREYNWVK 300
Db |||||
241 MSQAMIRGMDDEEGNQFVAYFLPVEETLKKRKRDQEEENDYAPDDVYDYKIAREYNWVK 300

QY 301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360
Db |||||
301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360

QY 361 LEAQEARKAQLNHEPEEEEEEMTEEEKAGSD-EEQEKSSSEKE 407
Db |||||
361 LEAQEARKAQLNHEPEEEEEEMTEEEKAGSD-EEQEKSSSEKE 407

RESULT 10
Q69N55 PRELIMINARY; PRT; 538 AA.
ID Q69N55
```


AC 09VNS5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE CG2503-PA (LD37523p).
 GN ORFNames=CG2503;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.B.,
 RA Abriel J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gonzalez J.H., Gu Z., Guan P., Harris M.,
 RA Hladik N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwa C.,
 RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong E., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537573;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 melanogaster euchromatic genome sequence.";
 RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review.";
 RN Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Lieb G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.J.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003605; AAF52095.1; -
 DR EMBL; AY070561; AAL48032.1; -
 DR FlyBase; FBgn0037274; CG2503.
 DR InterPro; IPR007133; Paf1.
 DR Pfam; PF03985; Paf1; 1.
 SQ SEQUENCE 538 AA; 60794 MW; D55E95B4F4EB8E51 CRC64;
 Query Match 45.0%; Score 1244.5; DB 2; Length 538;
 Best Local Similarity 50.0%; Pred. No. 3e-51;
 Matches 271; Conservative 66; Mismatches 172; Indels 38; Gaps 11;
 Qy 1 MAPIQTQAOQREDGHRPNSHRTLPERSGVVCRVKYCNLPDIPDPKFEITYPDONRFVQ 60
 Db 1 MPPTINNSAVNSAAEK-RPQOTERKSEIICRVKYGNNLPDIPDLKPLQPFDSHREVQ 59
 Qy 61 YKATSLKQKHDLITPDLDGVTIDLINPDYRIDPNVLLDPADKLEEEIOAPTSSKR 120
 Db 60 YNPTSLERNFKYDLVLTEDLGVTVDLINRELYQADSWTLDPADKLEESTLFTDSVR 119
 Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISN-EKPEVKIGVSVKQQTBEIYKDRDSQITAI 179
 Db 120 SRQHSRTVSWLRKSEYISTEQTRPQPONLEIAKVGYNVKKSLREETLYLDREAQIKAI 179
 Qy 180 EKTPEDAQSISQHSYKSPRTVPVEVPFDPFKMWINPCAQVIFDSDPAPKDTSGAAALE 239
 Db 180 EKTFSFDTKSIITHYKSPNVVPVEVLPFDPFTNWKFPQAVIFDSDPAPAKNVPAQLE 239
 Qy 240 MMSQAMIRGMWDEGNQFVAYFLPVEETLKKRKQDQSEEMDYAPDDVYDYKAREYNWV 299
 Db 240 EMSQAMIRGMWDESGEQFVAYFLPTEQTLKRTDFDINGELYKEEEVEYKAREYNWV 299
 Qy 300 KNKASKGYEENYFFIFREGDGVYNNLETRVRLSKRAKAGVQSGTQNALLVVKHRDMMNK 359
 Db 300 KTKASKGYEENYFFVFMQ-DGIYNELETRVRLNKRVRKVG-QQPNNTKLVYKHPRLDSM 357
 Qy 360 ELEQAEARKAQLNHEPEEP-----EEEM-----ETEE-----KEAGSGD----- 395
 Db 358 EHRMQRVRLQLEVPGESEEEIVEEVEEQMIIIGETEKTSDDAAGQAASGADSPAQV 417
 Qy 396 --EEQKSGSSSEKSGSEDEHSGSEEREEDGRDASDKSGSEDESEDESEDEARAAKDEEI 453
 Db 418 ARDQSRSRTRTSGS--SSGSGSGSGSRASRSKSGSRSGSRSTNSPAGSKSGSR- 475

QY 454 FGSDADSEDDADSDDEDRGOAOGSDNDSDGS-NGGQSRSHSRSSASPPFSGSEHSAQ 512
 Db 476 SRVSRRSRKSGSRSRKSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSGSA 535
 QY 513 ED 514
 Db 536 SD 537

RESULT 11
 Q7PXA3
 ID Q7PXA3 PRELIMINARY; PRT; 468 AA.
 AC Q7PXA3;
 DT 01-MAR-2004 (T-EMBLrel. 26, Created)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
 DE 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE AGCP12185 (Fragment).
 GN Names=agCG49165; ORFNames=ENSANGG00000017836;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAAB01008987; EAA01198.1; -.
 DR InterPro; IPR007133; Pfam1.
 DR Pfam; PF03985; Pfam1; 1.
 FT NON TER 468 468
 FT NON TER 468 468
 SQ SEQUENCE 468 AA; 53298 MW; 47BECB1BB520C826 CRC64;

Query Match 40.8%; Score 1129; DB 2; Length 468;
 Best Local Similarity 61.8%; Pred. No. 7.4e-46;
 Matches 223; Conservative 45; Mismatches 87; Indels 6; Gaps 5;

QY 1 MAPIQTQAOEDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFITYPFDQNRVQ 60
 Db 5 MAPIVQNGANGAD-KRP--VRQRRSELISRVKYNLTLPDIPDLKFITYPFDNRFIQ 61
 QY 61 YKATSLKQKHDLTPEDLGVITIDLPNTYRIDPNVLLDPADEKLDEEIQAPTSKR 120
 Db 62 YNPTSLENYRYEVLTEHDLGVITIDLPNTYRIDPNVLLDPADEKLDEEIHPTQDSMR 121
 QY 121 SQQKAVVPMWRKTEVISTEFNRYGISN-EKPEVKIGVSVKQOFTTEELYKORDSOITAI 179
 Db 122 SSRHAKSVSLRSEYISTEOTFNPTQWEKVEAKGVFNKSLREETLYMDREAQIKAI 181
 QY 180 EKTFFDAQKSIQHYSKRPVTPVEVMPVFPDFKWINPCAQVIFDSDPAKPDTSAGAALE 239
 Db 182 EKTFFEDTKPTTHYSKPGVTPVEIMPVFPDFANWKYPCAQVIFDSDPAPSGKNVPAQIE 241
 QY 240 MMSQAMIRGMMDSEGNQFVAYFLPVETLKKRKRDOEEMDYAPDDVYDKIAREYNWNV 299
 Db 242 EMSQAMIRGMVDSGEQFVAYFLPTDITLKKRRDLVNETLYDEDEEYEQWAREYNWNV 301
 QY 300 KMKASKGYEENYFFI PREGDGVVYNELETVRLSKRAKAGVSGTGNALLVVKHRDMNEK 359
 Db 302 KSKASKGYEENYVTLR-P-DGIYNELETVRLSKRRQK-NAQQQNTKLVLVVKHRPLNAS 359
 QY 360 E 360
 Db 360 E 360

RESULT 12
 P90783

ID P90783 PRELIMINARY; PRT; 425 AA.
 AC P90783;
 DT 01-MAY-1997 (T-REMLrel. 03, Created)
 DT 01-MAY-1999 (T-REMLrel. 10, Last sequence update)
 DT 01-JUN-2003 (T-REMLrel. 24, Last annotation update)
 DE Hypothetical protein C55A6.9.
 GN ORFNames=C55A6.9;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Kershaw J.K.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z81051; CAB02869.1; -.
 DR PIR; T20261; T20261.
 DR WormBase; WBGene00008338; C55A6.9.
 DR WormPep; C55A6.9; CR20614.
 DR InterPro; IPR007133; Pfam1.
 DR Pfam; PF03985; Pfam1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 425 AA; 49897 MW; 4BA06AF90913DCC5 CRC64;

Query Match 23.3%; Score 645; DB 2; Length 425;
 Best Local Similarity 33.1%; Pred. No. 5e-23;
 Matches 146; Conservative 96; Mismatches 165; Indels 34; Gaps 10;

QY 24 PERSGVCRVKYCNLSLPDIPDPKFITYPFDQNRVQKATSLKQKHDLTPEDLGV 82
 Db 14 PRKVDFMLKPRFTNTVPDPAKFWTCFFVPLGRFVFQPAAYRDYKGVICDDMDGL 73
 QY 83 TIDLNPITYRIDP-NVLLDPADEKLDEEIQAPTSKRQSOHAKVPMWRKTEVISTEF 141
 Db 74 NVDLIDLKKYDEDPITETIDEKONILLDDGAAKLIARSSQHSKLVPMWRKTEVISTEF 133
 QY 142 NRYGISNKEPKVIGVSVKQOFTTEELYKORDSOITAEKTFEDAKTSISCHYSKRVTP 201
 Db 134 NRGVTADROETKGLYNLKKQOVEDMYRDKSQSDAINKTFEDVRKPVKHYKGVKA 193
 QY 202 VEVMPVFPDFKWINPCAQVIFDSDPAKPDTSAGAALEMMSQAMIRGMMDSEGNQFVAYF 261
 Db 194 VEESVFPDFDHNKHLFAHVQFDGDTITTEFEEDERQOARESSVIKAMEREDQKFAAVF 253
 QY 262 LPVEETLKKRKRDOEEMDYAPDDVYDKIAREYNWNVKASKGYEENYFFI PREGDGV 321
 Db 254 VPTIGCLTSFMDLELERPFDEDMKYEFLSREYTFKMEHLPPR--DRDVFIMYHRNNVF 311
 QY 322 YNELETVRLSKRAKAGVSGTGNALLVVKHRDMNEKELEAQEAKQALNHE-PEEE 380
 Db 312 QYNEVCNVKMT-RKPRMALSRSKSLTLYR-----NPSELEQDMNKREAHLYEQKTRK 366
 QY 381 EEEME--TEEKAGSGDEEQEKSGSSEKSGSEDEHSGSESEEREEDRDSEASDKSGSEDE 438
 Db 367 QEILEKIQEKEEGD-----SSDQSSDSDDDKPKQSR-----SDSSSDV 406
 QY 439 SSEDARAARDKEEIFGSDAD 459
 Db 407 SSDDD--SPRKKEPTVSDSD 425

RESULT 13
 Q8RW91
 ID Q8RW91 PRELIMINARY; PRT; 589 AA.

Q8RW91.
01-JUN-2002 (TREMBlrel. 21, Created)
01-JUN-2002 (TREMBlrel. 21, Last sequence update)
05-JUL-2004 (TREMBlrel. 27, Last annotation update)
Atlg79730/F19K16.29.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
Hayashizaki Y., Ishida J., Jones T., Kaniya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan W.W.,
Ishida J., Jones T., Kaniya A., Karlin-Neumann G., Kawai J., Lam B.,
Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
[2]
DR EMBL; AY093780; AM103971; -;
DR EMBL; BT002711; AA011627.1; -;
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
DR SEQUENCE 589 AA; 67335 MW; 49D9F81P59C43AFF CRC64;
Query Match 13.5%; Score 372.5; DB 2; Length 589;
Best Local Similarity 25.5%; Pred. No. 5.4e-10;
Matches 127; Conservative 78; Mismatches 152; Indels 141; Gaps 20;
QY 6 QTOAQRED-GHR---PNSHRT-----LPERSGVVCRVKYCN 37
Db 143 ELEKQKQDEKHKRQOMKNSHKSQMPKGTBEKPTPLTTDRVQNRLLKPTTFICKLFRN 202
QY 38 SLDPDPDPKFTVPPQNRVVOYKATSLKQKHDLTTPDLGVTIDLNPYRIDPN 97
Db 203 ELDPSPQALKMTIKRDKQDFTKTITSLEKLWKPKIFVFPDGLGIPDLDDLVSYN-PPK 261
QY 98 VL--LDPADKLEEEIQAPTSRSQOQAKVVP-----WMRKTEYISTEFNRYGISNE 149
Db 262 VKAPLAPADEELLRDD--DAVTPIKDGIRRKERTDKGMSLWLVKTQYISS--INNE 314
QY 150 KPEVIGVSVKQOFTBEI-----YKDRDSQITAIKTFEDAKQSISQHY 194
Db 315 -----SARQSLTEKQAKELREMGGINILHNNRNRQIKDIEASF-ACKSRPVHA 365
QY 195 SKRPVTVPEVMPVPDPFKWNPINCAQVIFDS-----DPAPKDTSGAALEMMS 242
Db 366 TNKQLQVEVILPLPYFDRDEQVVFANFDGAPIADSEFFGKLDPSIRDHESRAL--LK 423
QY 243 QAMIRGMDBEGNQFVAYFLPVETLKKRKDOEEEMDYAPDDVYKYAREYNWVKNK 302
Db 424 SYVVAGSDTANPEKFLAYMVPSPDLSEKDIHDENEEISYT-----WVREYLDVQPN 475
QY 303 ASKGYENYFFIFREGDGVYNELETRVLSKRAKAGVQSG-----TNA 347
Db 476 AND--PCTVLVSPDNSTASYL-PLPMLNLKRRAREGRSSDEIHEFPVPSRVTVRRST 532
QY 348 LVLVVKHRD--MNEKLEAQARKAQLEN-----HEPEEBEEMETEKEAGSDSE 396
Db 533 VSVIEHKDSGVYSRVSASSKVRRLDEGLGRSWKHEPEQD----- 575

QY 397 EQKGSSESKEGSEDBHS 414
Db 576 -----ANQYSDGNEDDYS 588
RESULT 14
Q9MA04 PRELIMINARY; PRT; 593 AA.
AC Q9MA04;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE F20B17.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
Shinn P., Altafi H., Bei Q., Chin C., Chio J., Choi E., Conn L.,
Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
Ecker J.R.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
Khan S., Kim C., Altafi H., Bei Q., Chin C., Chio J., Choi E.,
Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
Theologis A., Ecker J.;
RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
DR EMBL; AC010793; AAF68118.1; -;
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
DR SEQUENCE 593 AA; 68056 MW; 3E188CF6A2P8E61A CRC64;
Query Match 13.1%; Score 361.5; DB 2; Length 593;
Best Local Similarity 25.8%; Pred. No. 1.8e-09;
Matches 130; Conservative 73; Mismatches 152; Indels 149; Gaps 21;
QY 6 QTOAQRED-GHR---PNSHRT-----LPERSGVVCRVKYCN 37
Db 143 ELEKQKQDEKHKRQOMKNSHKSQMPKGTBEKPTPLTTDRVQNRLLKPTTFICKLFRN 202
QY 38 SLDPDPDPKFTVPPQNRVVOYKATSLKQKHDLTTPDLGVTIDLNPDT 91
Db 203 ELDPSPQALKMTIKRDKQDFTKTITSLEKLWKPKIFVFPDGLGIPDLDDLVS 262
QY 92 YRIDPNVL--LDPADKLEEEIQAPTSRSQOQAKVVP-----WMRKTEYISTEFN 143
Db 263 YN-PPKAPLAPADEELLRDD--DAVTPIKDGIRRKERTDKGMSLWLVKTQYISS-- 316
QY 144 YGISNEKPEVKIGSVKQOFTBEI-----YKDRDSQITAIKTFEDAKQ 188
Db 317 --INNE-----SARQSLTEKQAKELREMGGINILHNNRNRQIKDIEASF-ACK 365
QY 189 SJLSQVSKRPVTVPEVMPVPDPFKWNPINCAQVIFDS-----DPAPKDTSGAA 236
Db 366 SPVTRATNKQLQVEVILPLPYFDRDEQVVFANFDGAPIADSEFFGKLDPSIRD--A 421
QY 237 ALEMMSQAMIRGMDBEGNQFVAYFLPVETLKKRKDOEEEMDYAPDDVYKYAREYN 296
Db 422 HESRVSYELPISMTANTANPEKFLAYMVPSPDLSEKDIHDENEEISYT-----WVREYL 473
QY 297 WNVKNKASKGYENYFFIFREGDGVYNELETRVLSKRAKAGVQSG----- 344

Db 474 WDVQPNAND--PGTYLVSPDNGTASYL-PLPMRLNLKRRAREGRSSDEIHFPVPSRV 530
Qy 345 ----TNALLVVKHSD--MNEKLELEAQEARKAQLN-----HPEPEEEEEEMETEKE 390
Db 531 VRRRSTVSVIEHKDGVSSRVGASSKWRLEDEGLGRSNKHEPEQD----- 579
Qy 391 AGSDDEQEKSGSSEKESGEDEHS 414
Db 580 -----ANQYSDGNEDDYS 592

RESULT 15

Q6ZD92 PRELIMINARY: PRT; 451 AA.
AC Q6ZD92;
DT 05-JUL-2004 (TREMELrel 27, Created)
DT 05-JUL-2004 (TREMELrel 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel 27, Last annotation update)
DE Proline-rich protein-like.
GN Name=P043808.25;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004460; BAC99509.1; -.
DR InterPro; IPR007133; Paf1.
DR Pfam; PF03985; Paf1; 1.
SQ SEQUENCE 451 AA; 52159 MW; 297158E4B56642AE CRC64;

Query Match 13.0%; Score 358.5; DB 2; Length 451;
Best Local Similarity 26.9%; Pred. No. 1.8e-09;
Matches 115; Conservative 80; Mismatches 180; Indels 53; Gaps 17;
Qy 17 PMSHR---TLPSRGVVCVKVCNSLPDIPDPKFTYFPDQNRVQVKATSLKQHKHD 73
Db 47 PNAERFENLKKPTPLCKHKEFNELDPSPSSQLKWLPLNKDKDRTYKVRITSLEKNYIPK 106
Qy 74 LITEPDLGVTIILNPDTYRIDP-NVLLDPADEKLL-EEIQAPTSS-----KRSQOHAKV 127
Db 107 MIVPEDLGITPLDLMVSNTFPVQPMAPDEDEELLRDDDEVLTVPKKGIRKERTDKG 166
Qy 128 VPMWRKTEYI---STEFNRYGI----SNEKPEVKIGSVKQOFTBEEIYKDRDSQITAE 180
Db 167 MSWLVTQYIISLSDAAKMSITEKQAKERREREG--RNTFLEN--INDREKQIKALE 221
Qy 181 KTFEDAQKISQHSYKRPVTPVEVMPVFPDFQXWINPCAQVIFDSDPAPKDTSGAAALE- 239
Db 222 DGFPR-AAKSRPVHOTKRGMEAEWVLPPLPDFORYDDQFVMVNFQGDPT-ADSEQYNKLER 279
Qy 240 -----MMSQAMIRGMDEEGNQFVAYFLPVEETLKRRKQDEEMDYAPDDVVDY 289
Db 280 SERDECESAVMKGSFLVNGSDPAKQEKFLAYMVPSHELKSLDDEDEDIQYS----- 332
Qy 290 KIAREYNMNVKNAKSGYEENFFIFREGDGVVYNELETRVRLSKRAKAGVQSGTNALL 349
Db 333 -WLREYHWEVRGD-DKDDPTTYLVTF-DDDGAKYLPPLTKVLQKKKKEG-RSGDE--- 385
Qy 350 VVKHRDMNKELEAQEARKAQLNHEPEEEEEEMETEKEAGGSDDEQEKSGSSEKES 409
Db 386 -IEHFPVPSRITVSRTHAGGMMEHGESSNMH-----ENLKRQRSSVDDDLIDHPKHRSVED 440
Qy 410 EDEHSGSE 417
Db 441 MDQYSGDE 448

Search completed: September 26, 2005, 13:54:48
Job time : 183 secs